

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2000, 03:02:45 ; Search time 2798.97 Seconds  
(without alignments)  
893.952 Million cell updates/sec

Title: US-09-246-451-1

Perfect score: 1402  
Sequence: 1 ctgcagatcgttaccgct.....gcactacacaaagcggtata 1402

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 972840 seqs, 892348106 residues

Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_om:\*  
5: gb\_om:\*  
6: gb\_ph:\*  
7: gb\_ph1:\*  
8: gb\_ph12:\*  
9: gb\_ph12:\*  
10: gb\_ph12:\*  
11: gb\_ph13:\*  
12: gb\_ph13:\*  
13: gb\_ph13:\*  
14: gb\_ph13:\*  
15: gb\_ph13:\*  
16: gb\_ph13:\*  
17: gb\_ph13:\*  
18: gb\_ph13:\*  
19: gb\_ph13:\*  
20: gb\_ph13:\*  
21: gb\_ph13:\*  
22: gb\_ph13:\*  
23: gb\_ph13:\*  
24: gb\_ph13:\*  
25: gb\_ph13:\*  
26: gb\_ph13:\*  
27: gb\_ph13:\*  
28: gb\_ph13:\*  
29: gb\_ph13:\*  
30: gb\_ph13:\*  
31: gb\_ph13:\*  
32: gb\_ph13:\*  
33: gb\_ph13:\*  
34: gb\_ph13:\*  
35: gb\_ph13:\*  
36: gb\_ph13:\*  
37: gb\_ph13:\*  
38: gb\_ph13:\*  
39: gb\_ph13:\*  
40: gb\_ph13:\*  
41: gb\_ph13:\*  
42: gb\_ph13:\*  
43: gb\_ph13:\*

44: gb\_hcg7:\*  
45: em\_hcg1:\*  
46: em\_hcg2:\*  
47: em\_hcg3:\*  
48: em\_hcg4:\*  
49: em\_hcg5:\*  
50: gb\_ph13:\*  
51: gb\_ph13:\*  
52: gb\_ph13:\*  
53: gb\_ph13:\*  
54: gb\_ph13:\*  
55: gb\_ph13:\*  
56: gb\_ph13:\*  
57: gb\_ph13:\*  
58: gb\_ph13:\*  
59: gb\_ph13:\*  
60: gb\_ph13:\*  
61: gb\_ph13:\*  
62: em\_hcg4:\*  
63: em\_hcg5:\*  
64: em\_hcg6:\*  
65: em\_hcg7:\*  
66: em\_hcg8:\*  
67: gb\_hcg18:\*  
68: gb\_hcg19:\*  
69: gb\_hcg20:\*  
70: gb\_hcg21:\*  
71: gb\_hcg22:\*  
72: gb\_hcg23:\*  
73: gb\_hcg24:\*  
74: gb\_hcg25:\*  
75: gb\_hcg26:\*  
76: gb\_hcg27:\*  
77: gb\_hcg28:\*  
78: gb\_hcg29:\*  
79: gb\_hcg30:\*  
80: gb\_hcg31:\*  
81: gb\_v11:\*  
82: gb\_v12:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1402	100.0	1578	1 PSECCAMC	M12546 Pseudomonas
2	282	20.1	2606	1 PSECCAMC	D14680 Pseudomonas
3	140.8	10.0	10057	2 AE000101	AE000101 Rhizobium
4	140.8	10.0	34010	1 RSPNGR234	Z68203 Rhizobium s
5	140.8	10.0	236165	5 A79351	A79351 Sequence 2
6	140.8	10.0	236165	5 A93003	A93003 Sequence 2
7	112.6	8.0	3078	1 SERCP450A	M83110 Saccharopol
8	110.6	7.9	5833	2 AF147703	AF147703 Streptomy
9	107.4	7.7	12905	2 AF055922	AF055922 Streptomy
10	104.2	7.4	63734	2 AF127374	AF127374 Streptomy
11	103	7.3	10167	2 SFU08223	U08223 Streptomyce
12	98.2	7.0	2500	2 SSU65940	SSU65940 Streptomyce
13	97.6	7.0	1233	5 E06907	E06907 DNA encodin
14	97.6	7.0	1274	1 SFMCP450	D30815 Streptomyce
15	92	6.6	1950	1 PSECCAMBA	D00528 Pseudomonas
16	90.4	6.4	7870	1 STE18574	Y18574 Streptomyce
17	89.2	6.4	2218	2 STE250199	AJ250199 Streptomy
18	88.4	6.3	36030	1 MTCY50	Z77137 Mycobacteri
19	88.2	6.3	1629	1 STMSUACB	M32238 S.griseolus
20	88.2	6.3	22574	14 ABO18074	Y18556 Cloning vec
21	87.4	6.2	2795	1 AB018074	AB018074 Streptomy
22	87	6.2	1400	1 STWOLEP	L37200 Streptomyce
23	83.4	5.9	1194	5 A48328	A48328 Sequence 6
24	83.4	5.9	4496	5 A48326	A48326 Sequence 4

C	25	82	5.8	39524	1	SCH10	AL049754	Streptomy
C	26	81.6	5.8	6937	2	AF040571	AF040571	Amycolato
C	27	81.6	5.8	107379	1	SHGCP1R	X86780	S.hydroscop
C	28	80	5.7	3104	1	STMCHOPA	M31939	Streptomyce
C	29	78.2	5.6	8366	2	AF072709	AF072709	Streptomy
C	30	78	5.6	9882	1	AMOXVAB	Y16952	Amycolatops
C	31	77.8	5.5	53784	1	AMM23012	AJ223012	Amycolato
C	32	77.8	5.5	53789	5	A69720	A69720	Sequence 3
C	33	77.8	5.5	76199	2	AF040570	AF040570	Saccharopol
C	34	77.4	5.5	8407	2	SEU82823	U82823	Saccharopol
C	35	77.2	5.5	42861	1	MPCY339	277163	Mycobacteri
C	36	77.2	5.5	107379	1	SHGCP1R	X86780	S.hydroscop
C	37	76.6	5.5	2168	1	MPDUMCG	D16098	Micromosp
C	38	76.6	5.5	2168	5	E08092	E08092	Genomic DNA
C	39	74	5.3	1470	2	AF087022	AF087022	Streptomy
C	40	73.6	5.2	1700	1	SGSOYBC	X63601	S.griseus S
C	41	73.6	5.2	1735	5	I15434	I15434	Sequence 10
C	42	73.6	5.2	4342	2	AF079139	AF079139	Streptomy
C	43	73	5.2	47852	1	MTV023	AL022022	Mycobacte
C	44	71.8	5.1	2243	2	SERERYFGH	M54983	Saccharopol
C	45	71.4	5.1	1688	1	STMSUBCB	M32239	S.griseolus

ALIGNMENTS

RESULT	1
LOCUS	PSECAMC
DEFINITION	Pseudomonas putida camc gene, encoding cytochrome P-450-cam, complete cds and cytochrome P-450-camA gene, 5' end.
ACCESSION	M12546
VERSION	M12546.1 GI:151114
KEYWORDS	cytochrome P450cam,
SOURCE	P.putida (ATCC 17453) DNA, clone pK6300.
ORGANISM	Pseudomonas putida
	Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;

REFERENCE	1 (bases 1 to 1578)
AUTHORS	Unger,B.P., Gunsalus,I.C. and Sliagar,S.G.
TITLE	Nucleotide sequence of the pseudomonas putida cytochrome P-450-cam gene and its expression in Escherichia coli
JOURNAL	J. Biol. Chem. 261, 1158-1163 (1986)
MEDLINE	86111751
FEATURES	Location/Qualifiers

source	1..1578
	/organism="Pseudomonas putida"
	/db_xref="taxon:303"
	156..1403
	/note="cytochrome P-450-cam"
	/codon_start=1
	/transl_table=11
	/protein_id="AAA25760.1"

CDs	/db_xref="GI:151115"
	/translation="MTETITQSNANLAPPHVPEHLVDFEDMYNSNL.SAGVQEMAWA VLOESNVPLIVTRNGHMIAIRGOLIEAYDHHFSECFIRREGAEVADFTPT SMDPPEORFALANOVGMPPVVDKLENRIOELACLSIESLRGOGCNTEADAEFP IRIFMLAGLPEDIDPHLYKLTIDQMTRPAGSMFAFAKELVLYLPIIEOROKGT DAISTYANOVNRPITSDPAKRMCCLLVGCUDTYVNNLSPEMELASPERQELI ERPERIPANCEBELRRFSLVADGRILITSDERHGVOLKGDQILLFQMSGLDERENA CPMHVDSROKXSHTEFGHSHLCIOHLARREIIVTLKEMILRIDPISAPGAIOIH KSGIVGVALPMDPATTKAV"
	1425..>1578
	/note="cytochrome P-450-camA (gfg start codon)"
	/codon_start=1
	/transl_table=11
	/protein_id="AAA25761.1"
	/db_xref="GI:151924"
	/translation="MNANDNVIVGTGLAGVEAFGLRASGMEGNIIRLVGDATVIPHH LPLPSKA"

BASE COUNT	350 a	480 C	436 g	302 t
ORIGIN	5 bp upstream of PstI site.			

Query Match 100.0%; Score 1402; DB 1; Length 1578;  
Best Local Similarity 100.0%; Pred. No. 2,9e+283;  
Matches 1402; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	ctcgaagatcgttatccgcctgcccgaatcgtatcaccacgcttltccatcgacgagcc	60
Db	1	CTCAGAGATCGTTATCCGCGCGGATGATCACCAGCGTTTTCATGACGAGAGCC	60
Qy	61	agcaagcacttaactgacgtgcaagagcagcgtatcaaacccgtgacgcactcact	120
Db	61	AGCAAGGACTTAACCTGTCAGAGCGAGACCTGATCAACCCGATGACTCCACT	120
Qy	121	ctttagcaaccgcgcttccagagaaacaacaatgagacttgaacaacatacaagca	180
Db	121	CTTTAGCGAACCCGCGCTTCCAGGAGAAACAACAATGAGACTGAAACATCAAA	180
Qy	181	acgcacactcttgcctcctcgtccaccccaatggtccagagcaccctgtaattcga	240
Db	181	ACGCCAATCTTGCCCTCTGACACCCCATGTGCGAGACACTGTGTGATCGAC	240
Qy	241	tgtacaaatccgttgatctgtctgcccgggtgacagagccttgagcgttctcaga	300
Db	241	TGTAAATCCGTGGAATCTGTGCGCGCTGCGAGAGGCTTGCGAGTTCTCA	300
Qy	301	caaacgtaccggaatctggtgagctcgtcgaacgacgacacgtgacacacacgc	360
Db	301	CAAAAGTACCGGATCTGTGTGACTGTGCAAGGCGGACATGATGCCACTCGC	360
Qy	361	gcaactgacgtcgtgagcgtatgaagatcaccgcaacttccagcagatgcgcgtca	420
Db	361	GCCAACTGATCTGTGAGCGCTATGAAATACCGCCACTTTTCAGCGATGCCG	420
Qy	421	tcctcgtgaaagcgggaaagcctacgacttaccacacttgatgagtcggccgagc	480
Db	421	TCCCTGTGAAGCGGCGAAGACCTACGACTTATCCACCTGATGATGCCCGAG	480
Qy	481	agcgcacgttctgctgctgccaaccaaagtgtgtgcaacgacggtggtgataa	540
Db	481	AGGCGCGTTTTCGTCGCGTGGCAACCAAGTGTGTGCAAGCGGATGATGAG	540
Qy	541	agaacggatcagcagagcgtgctgtcgtgatcgagagcgtgcgcgcaagacag	600
Db	541	AGAACGGATTCAGAGACTGCGCTGCTGATCGAGACCTGCGCCCAAGACAG	600
Qy	601	gcaactcaccgaggaactacgcaacccttccgatacgcattcattcgtctgcag	660
Db	601	GCAACTCACCGAGAGACTACCGAACCTTCCGATACGATCTTCATCTGCTG	660
Qy	661	gtclaccggaagaagatattccgcaacttgaatacctaaagatcagatgacccgtc	720
Db	661	GTCACCGGAAGAATATCCCGCACTTGAATACCAACGATCAACCCGCTCGG	720
Qy	721	atcgacgacatgaccttcgcaagagcgaagagcgctctacgacatctgataccga	780
Db	721	ATCGACACATGACCTTCGAGAGGCGAAGGCGGCTCTACGATCTGATACCG	780
Qy	781	tcgaagcagcagcaggaacccggaacccgactacacatcgtttgccaagcgag	840
Db	781	TCGAGCAACCGAGCAAGCCGGAACCGACCTATCACATCTGTCACAGGCCAG	840
Qy	841	tcaatggcgacacatcaccagtgacgaacccaagagatggtgtgaccttactgctg	900
Db	841	TCAATGGCGACACATACCAGTACGACCAACCAAGAGATGTGTGGCTTACTGT	900
Qy	901	gcggccttgatagcgttgatcaatttccctcagcttcagcatgagatcttcctg	960
Db	901	GCGGCGCTGATAGCGTGTGATTTCTCCTACGTTACGATGAGATCTTCTG	960
Qy	961	cggagcatcgccagagcgtatcagaggtccagagcgtatccagcgcgttggaag	1020
Db	961	CGGAGCATCGCCAGAGCTGATGAGCGTCCGAGCGTATTCACCGCGCTTGAG	1020

Qy 1021 tactccgagccttcgctggttgccgagtcgacccctaccctccgattacagtttc 1080  
|||||  
Db 1021 TACTCCGAGCCTTCTCGCTGGTGGCCGATGCGCATCTCCACTCCGATTAGAGTTTC 1080  
Qy 1081 atggcgctgcaactggaagaagtgaccagatcctgtctaccgagatgctgtcgtg 1140  
|||||  
Db 1081 ATGGCGGCACTGAAAGAAAGTGACCGATCTCGTACCGGAGATGCTGTCTGGCTGG 1140  
Qy 1141 atgagcgcgaagaacgcttcgacgagtcgacgttcgacctcagtcgcaaaagtctcacaca 1200  
|||||  
Db 1141 ATGAGCGCCGAAAGACGCTGCCCGATGACGTCGACTTCAGTCGCCAAAGGTTTCACACACA 1200  
Qy 1201 ccacacttggccacgagcagcaccatctgtcccttggccagcaccctggcccgaggaatca 1260  
|||||  
Db 1201 CCACCTTTGGCCACGCGGACATCTGTGCTGGCCAGCAACTGCGCCGCGGAATCA 1260  
Qy 1261 tcgtacccctcaaggaatgctgacgagatctcctgacttcctcatctgcccgggtgcc 1320  
|||||  
Db 1261 TCGTACCCCTCAAGGAATGCTGACGAGATCTCTGACTTCTCCATTTGCCCGGGTGCC 1320  
Qy 1321 agattcgcacgaagagcgagcagtcgacgagcggtgacgagcaccctctgtctgagtc 1380  
|||||  
Db 1321 AGATTGCGCACAGAGCGGATGCTGACGCGGTCGACGAGCATCTCTGTGTTGGGATC 1380  
Qy 1381 cggcgaactacaaagcggtata 1402  
|||||  
Db 1381 CGGCGACTACCAAGCGGTATA 1402  
RESULT 2  
PSECAMD 2606 bp DNA BCT 04-FEB-1999  
LOCUS pseudomonas putida plasmid CAM cam repressor (camr) gene and  
DEFINITION 5-exo-hydroxycamphor dehydrogenase (camd) gene, complete cds.  
ACCESSION D14680 D14452 D14453  
VERSION D14680.1 GI:473744  
KEYWORDS 5-exo-hydroxycamphor dehydrogenase; cam repressor; cytochrome  
P-450cam hydroxylase operon repressor.  
SOURCE pseudomonas putida (strain:Pg1) plasmid:CAM DNA, clone:pp1.  
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;  
Pseudomonas.  
REFERENCE 1 (bases 1 to 2606)  
AUTHORS Aramaki,H.  
JOURNAL Direct Submission  
TITLE Submitted (17-FEB-1993) to the DDBJ/EMBL/GenBank databases.  
Hironori Aramaki, Daiichi College of Pharmaceutical Sciences,  
Department of Microbiology; Yamagawa 22-1, Minami-Ku, Fukuoka,  
Fukuoka 815, Japan (Tel:092-541-0161(ex.415), Fax:092-553-5698)  
2 (sites)  
Koga,H., Aramaki,H., Yamaguchi,E., Takeuchi,K., Horiuchi,T. and  
Gunsalus,I.C.  
TITLE camr, a negative regulator locus of the cytochrome P-450cam  
hydroxylase operon  
J. Bacteriol. 166 (3), 1089-1095 (1986)  
86223770  
3 (sites)  
Aramaki,H., Koga,H., Sagara,Y., Hosoi,M. and Horiuchi,T.  
TITLE Complete nucleotide sequence of the 5-exo-hydroxycamphor  
dehydrogenase gene on the CAM plasmid of Pseudomonas putida (ATCC  
17453)  
JOURNAL Biochim. Biophys. Acta 1174 (1), 91-94 (1993)  
MEDLINE 93326643  
4 (sites)  
Aramaki,H., Takeuchi,K., Koga,H., Sagara,Y. and Horiuchi,T.  
TITLE Nucleotide sequence of the gene encoding the repressor for the  
cytochrome P-450cam hydroxylase operon on the CAM plasmid of  
Pseudomonas putida Pg1  
unpublished (1993)  
JOURNAL 5 (bases 1 to 2606)  
REFERENCE Aramaki,H., Sagara,Y., Takeuchi,K., Koga,H. and Horiuchi,T.  
AUTHORS Nucleotide sequence of the gene encoding a repressor for the  
TITLE cytochrome P-450cam hydroxylase operon on the Pseudomonas putida

CAM plasmid  
JOURNAL Biochimie 76 (1), 63-70 (1994)  
MEDLINE 94304997  
FEATURES  
source location/Qualifiers  
1. 2606  
/organism="Pseudomonas putida"  
/plasmid="plasmid CAM"  
/strain="Pg1"  
/db\_xref="taxon:303"  
/clone="p1p1"  
/complement(482, .1042)  
/gene="camr"  
complement(482, .1042)  
/gene="camr"  
/function="a negative regulator"  
/codon\_start=1  
/transl\_table=11  
/product="cam repressor"  
/protein\_id="BAA03510.1"  
/db\_xref="GI:473745"  
/translation="MDIKOSLIHAMRLISAKRGDQATMPICAEVGYTPPTIYHNYG  
DIQGLIKRAIDETRYOVAEATYHGTBERGPKIGTRGMATFIQFATSEPMCMVLVOH  
IMAGEPSSWVADTLRGVADLQFHAQGRLTFFPREBAOLIMMGALGALLYALSRGA  
GYQDLALQKAKIDITVLAFFNIEE"  
1239, .1244  
-35\_signal 1262, .1267  
-10\_signal 1355, .1362  
RBS 1365, .2450  
gene /gene="camd"  
1365, .2450  
CDS /gene="camd"  
/codon\_start=1  
/transl\_table=11  
/product="5-exo-hydroxycamphor dehydrogenase"  
/protein\_id="BAA03511.1"  
/db\_xref="GI:473746"  
/translation="MOYARAARVAWEQNRVETWEPPIEDPAPGALRVVVLGVCGSDV  
HYSGRGAMPRIITIGHGIGIKETIGYTTDVAGVPRKODMYMARIALCHCH  
SCVIDETPRDNTSFFEHAKRPWGSYADPACLPKMATFRLPDHAKPELALAGCAL  
PVLKGYDRGVLDDTVVQAGVLAALVLAASGAKDITAIIDHSPTIDMARS  
LGATFETISLADTTPREERORLVDERFGRGSLVAAGALPAPREGVNLGNHGRVYI  
LGLMGAIGTQIPISPRDLITIKNMSIAGATFPRKHYIQAMQLAARLDRLADLITQR  
FSIDEASKALELVKAGALIKRPVIDSTL"  
1654  
old\_sequence  
BASE COUNT 535 a 830 c 723 g 518 t  
ORIGIN  
Query Match 20.1%; Score 282; DB 1; Length 2606;  
Best Local Similarity 100.0%; Pred. No. 3e-49;  
Matches 282; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ctgagagatcgtatccgcgtggccgacatcgtatcacccaggttttccatcgacgaagcc 60  
|||||  
Db 2325 CTGAGGATCGTTATCCGCTGGCCGATCTGATCACCCAGCGTTTTCATCGAGAGGCC 2384  
Qy 61 agcaaggaacttgaactggttcaaggcaggaagcaggaactgtatcaaacccgtgacgaactc 120  
|||||  
Db 2385 AGCAAGCACTTGAACTGGTCAAGGACGAGGACTGATCAAAACCCGTGATGACTCCACT 2444  
Qy 121 cttagccaaccgcgttccaaggaagaacaacaactgaactgtgaatacaataaagca 180  
|||||  
Db 2445 CTTTAGCCAAACCGCGTTCAGAGGAGAACAAACATGAGAGACTGAAACCATACAAAGCA 2504  
Qy 181 acgccaacttggccctctgcaaccccatgtgccaaggaacacgtgtatcgaacttgaca 240  
|||||  
Db 2505 ACGCAATCTTGCCCTCTCCACCCCATGTGCCAGAGCACTGTGATTCGACTTCGACA 2564  
Qy 241 tgtacaatcgtgaactctgtccggcggtgcaagagacct 282  
|||||

Db 2565 TGACAAATCCGTCGATCTGTGCGCGGCTGCAGAGGCCCT 2606

RESULT 3  
AE000101  
LOCUS Rhizobium sp. NGR234 plasmid pNGR234a, section 38 of 46 of the complete plasmid sequence.  
DEFINITION  
AE000101 000090  
VERSION AE000101.1 GI:2182671  
KEYWORDS  
ORGANISM Rhizobium sp. NGR234.  
SOURCE Rhizobium sp. NGR234.  
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Rhizobiaceae; Rhizobium.  
REFERENCE 1 (bases 1 to 10057)  
Freiberg, C., Fellay, R., Bairoch, A., Broughton, W.J., Rosenthal, A. and Perret, X.  
TITLE Molecular basis of symbiosis between Rhizobium and legumes  
JOURNAL Nature 387 (6631), 394-401 (1997)  
MEDLINE 97305956  
REFERENCE 2 (bases 1 to 10057)  
Freiberg, C.  
TITLE Direct Submission  
JOURNAL Submitted (22-NOV-1996) Genome Analysis, Institute for Molecular Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany  
3 (bases 1 to 10057)  
Freiberg, C.  
TITLE Direct Submission  
JOURNAL Submitted (12-DEC-1997) Genome Analysis, Institute for Molecular Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany  
update by submitter

REMARK  
FEATURES  
location/Qualifiers  
1. 10057  
/organism="Rhizobium sp. NGR234"  
/plasmid="pNGR234a"  
/strain="NGR234"  
/db\_xref="taxon:394"  
<1. 8279  
/organism="Rhizobium sp. NGR234"  
/db\_xref="taxon:394"  
/clone="PXB286"  
complement(206..254)  
/note="nod box-like sequence"  
/note="Region: nod box 13"  
/evidence=not\_experimental  
complement(418..984)  
/gene="y4vD"  
complement(418..984)  
/gene="y4vD"  
/note="hypothetical 21.4 kd protein possibly participating in a redox reaction"  
/codon\_start=1  
/transl\_table=11  
/product="y4vD"  
/protein\_id="AAB91892.1"  
/db\_xref="GI:2182672"  
/translation="MPVKRRVPVAFRTVRDETTGGPNPYRMEVTTEDYSGRKRV LFSLPGFPTFCSTQOLPDPERLVDEFKVGIVAVYCLSDAFVNNAMKALGKRV RLIPDSGETRKMGLVAKDNLGFGMSRMYAAVYNDSVGEQMEEFESDNCEDP VYASSPQNTLETRTPDTRARIGVPIKF"  
complement(1088..1102)  
/note="putative sigma-54 dependent promoter (-12/-24 region)"  
/note="Region: sigma-54 promoter 10"  
/evidence=not\_experimental  
1214..1229  
/note="putative NifA-dependent upstream activator sequence"  
/note="Region: UAS 10"  
/evidence=not\_experimental  
1242..1257  
/note="putative NifA-dependent upstream activator

misc\_feature  
sequence"  
/note="Region: UAS 11"  
/evidence=not\_experimental  
1346..1404  
/note="putative sigma-54 dependent promoter (three possibilities for -12/-24 regions with minimal consensus sequence [GG-N(10)-GC])"  
/note="Region: sigma-54 promoter 11"  
/evidence=not\_experimental  
1442..2134  
/gene="nifQ"  
/gene="nifQ"  
1442..2134  
/gene="nifQ"  
/note="NifQ protein homolog (possibly involved in molybdenum processing)"  
/codon\_start=1  
/transl\_table=11  
/product="NifQ"  
/protein\_id="AAB91893.1"  
/db\_xref="GI:2182673"  
/translation="MSNLAQVRAALSKGRVTLGIRMTDRPGMRQLSLDILGPMPTDL EMDPDQVFAVLSRALREIDAGCATATPATGLSOVELDITNRSRPTIVFRLER VVDSERGPPEALLRGLLHAHARAGDSASVLPFKITARRALRHDHRLGLDREVELS RLDAHFPTLAAQNTQNMKRRFYKRLCEABGYSICTAPSCPCKELESCEGSEGE SASPVGRVGGEA"  
2193..3707  
/gene="dctA1"  
2193..3707  
/gene="dctA1"  
/note="C4-dicarboxylate transport protein from the symbiotic plasmid; chromosomal copy exists; significant difference to acc. no. S38912"  
/codon\_start=1  
/transl\_table=11  
/product="dctA1"  
/protein\_id="AAB91894.1"  
/db\_xref="GI:2182674"  
/translation="MRGLRVCMHVEEIIILIVENLAENVKTPHYRHLVQVLAIAV GILGIFYPDGSKMKPLGDAFLIMVKMIIAIVIFLTVATGIAGMTDLAKVGRVAGKA MIYFIFLSTALLGLGVANVAVVOPGAGMIDPASAIDAKIATYAKAHQSGTGLMN IIPFTLVGAFAGSDILOVIFISVLRGISLAIVKKAENAVYDHLALDITFLVATILM KAPIGAFGAMAFITGKYGVASIANLAMLIGTFYLTSLFVPMVLGAARVNGFSIVA LIRYKEILLVLGTSSEALPGLNMKMGKRSVGLVPIPGVSFNLGNTIYMA LTAALFIAQATDTPISYGDQIILLILAMLSKGAAGITGAGFTTLAATLSAVSPVA GQALILGIDRPFSECRATINIIGNAAVATTIVAKMEGELPAQDLATTIAGKAVPEMVG LSSQSDVLELQGVKVLFGATNSADRLAGRPGGRSRRITAPDHSNQVPGPILSL"  
2240..>10057  
/organism="Rhizobium sp. NGR234"  
/db\_xref="taxon:394"  
/clone="PXB368"  
3704..4948  
/gene="y4vG"  
3704..4948  
/gene="y4vG"  
/note="putative cytochrome P-450"  
/codon\_start=1  
/transl\_table=11  
/product="y4vG"  
/protein\_id="AAB91895.1"  
/db\_xref="GI:2182675"  
/translation="MSDLRRKRVKTNPIPDHYPPALVYHFSLFTSPGNAAPPNGOPHA AVACHVDGPPPIFYSPSNTRDGRGWVITRADQRRVLEDTFTSSHSRIFASALGEH MWIVIPLEDDPPAHGVRFLNLPFSSRRVLALEPTIHAARGAALDICAIEKSCDVMK DPALPFTFVSFLGLSQRSESVLGVNSDLHGNAERRAARSVAVIDENAMAR RKSPLVDENMTFVVOAKTEGRSLJEEVRIGVLFIYAGIDPVYAAAIGDMYALANPK HOELLNREPARGLAAEELLRAYSTQIIRVATKRIEEGSGVIRBGDIYSCPAMIANR DTSERCEPTIDILARODNCHTAFGTGPHLCGHAAHARREIVIGLEWMLARIPAFRIKE GTAPITHGCHVGISNIIITWA"  
4949..5605  
/gene="y4vH"  
4949..5605  
/gene="y4vH"  
/note="hypothetical 24.6 kd protein"



		/codon_start=1
		/transl_table=11
		/product="Y4VH"
		/protein_id="AAB91896.1"
		/db_xref="GI:2182676"
		/translation="MDGRGSSNEAPIAPRSSHSTLTASHRTMNEESTINLPDREAIPDC LHYRCRGITDADEALBSCYPDPARDNPGSSSGSAAFTQHALAVETKRNRIHOLTN IILIEFGATCAVAEVSFTIALORGHDAIGELIKOYLGLCRVCDLPFKRGEMIVERTVW YDWVEKQRPKRSSEAEFRGPPIAGPHPDVPELYLSKRVAANAHPKDLIGSGEEP "
gene	5662..7308	/gene="y4vi"
CDS	5662..7308	/note="short chain type dehydrogenase/reductase"
		/gene="y4vi"
		/note="short chain type dehydrogenase/reductase"
		/codon_start=1
		/transl_table=11
		/product="Y4VI"
		/protein_id="AAB91897.1"
		/db_xref="GI:2182677"
		/translation="MDROYOPGRFVVVTGAAGIGRALVDIFANGDVVAVDLPDS GVTEIANGGEPLHGLEVDYSDREDDVALBALBKERSRLIEVLNNNGIGTPMACTAD TALEDROALAINVGASVACERPAKIMKGAAIVNAISGLLGNGKSANVASAKG LISTSKSLACRAMASRGIRIVTAVAPGHRTPMVAEPLAKIDSAITRRPPLGIARIAP DETARAVFLASAOASTITGSTLVVDGGMSVNOPGGAHOADPTPAEFMRPEDTV ARVYVLFEGATGVGGAAILRRFAENGDFVYIADGSEPAVKLAGILDKHLSRRVDRTV ETEVSYSFEELREBFGLDYEVFNMGMEILVPNTESSEPVLRILDMVLGTAFVCBEAA AISRRGSVILNLGASITSLSPLASSHAYGANAGIDMLTRCTAAELGPIRTVAPAD GYTRTCANRLAAVACGDMSAIStRORIPIGLRVGDAEEVAAEAFLASFDASYINSILHH VDGLISSRPAGMSCFVDGAI SAEMRQRRRAAMRWLLSP"
gene	7446..8501	/gene="y4vj"
CDS	7446..8501	/gene="y4vj"
		/note="hypothetical 39.2 kD protein; distantly related to albanal monooxygenase alpha chains (bacterial luciferases)"
		/codon_start=1
		/transl_table=11
		/product="Y4VJ"
		/protein_id="AAB91898.1"
		/db_xref="GI:2182678"
		/translation="MEPATFIILAQRGT HQSSASVYINSTEQALISDQAGSTRAMPRE HHPNNYSILVPSPLMVAHFACGLSTIRLGAVCVLPYOPQLLAETGFVDVANGLRL ELGSGSQOEFEEDRFRCNVIDEAPAIRSECIDLILKKGIETHSGRYMOIPTAISL VRTLOKPTPIPMIAITASKTMARAYVRGNHLFWYALHDGETTGILGIIKTAGSBG KEVDRSVKSLIRCCYASDDGAEINSYIDNNRFQRLSEAQQRRQSKDGIMEEMPT GDIISPTMKRNKLPIGSIINRVIRLLEIIDLKPODALOTQSDPFQCKTMLROIETLM GDITIPVOKSLGSEA"
misc_feature	8887..8904	/note="putative NifA-dependent upstream activator sequence"
		/note="Region: UAS 12"
		/evidence="not_experimental"
		8936..9010
misc_feature		/note="putative sigma-54 dependent promoter (-12/-24 region)"
		/note="Region: sigma-54 promoter 12"
		/evidence="not_experimental"
Query Match	10.0%; Score 140.8; DB 2; Length 10057;	
Best Local Similarity	48.5%; Pred. No. 8.5e-20;	
Matches 478; Conservative	0; Mismatches 502; Indels 6; Gaps 3;	
OY	324 actgcctgcaacgcgcgaactgatgcgccactcgcgcgccaactgatccgtcgagccat	383
DB	3890 ACGGCCTGATGACGAGGACTCGGTGATTACACGCCGGCAGAGACCACGCCGGCTCCTT	3949
OY	384 gaagattaccgcacatttcacgacgagggcccgttcatccctctgtaagccgacgaagcc	443
DB	3950 GAGGATTACAGAGACCTTTTCCACGCATTCGACGACATCTTCCCTCGGCACTTGCGCACAC	4009
OY	444 tacgaattcattccacactgatgatccgcgcgcgacgacgacgacagtcttgtagctgtgccc	503

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
Db	4010	TTGCCGGTCATCCCCCTTCGAAATTGGACATCCCGCCCATATGAGACTGTTTGTCTCACTGCTA	4069			
Oy	504	aaccaagtggttgcacatgcccggctggttgatgaataagcaacggaatccgaagagctggcc	563			
Db	4070	AATCTCTGTTCTCATATCCAGCGGGGTGTGGCGCTGGAGCCGACTATTCATCCACAGACA	4129			
Oy	564	tgtctgcgtatgcagaagcct--gcgcgcgcaagagcagtgcaacttccacggaagtaac	620			
Db	4130	GGCGCGGTGATCGACTCATTTGGCAAGAAAGAACACAGCTGTACGTGATGAAGGATTTC	4189			
Oy	621	ggcgaaacctcccgatacgaatcttcaatgctgctgcgaagttctacggaagaatatac	680			
Db	4190	GCTTGGCCCTTTACGTTTCACAGCTTTTTCCT--CACCTTTCTGGGACTTTCCCAAGCGCATC	4248			
Oy	681	ccgacttgaataatcaacggaatcagaatgaaaccggtccggaatgagcaatgacattgcga	740			
Db	4249	CGAAGTACTTGTCCGGCTGGGTGAACGATTTGGTTCACGGGACGCAAGAAACGGAAGC--	4306			
Oy	741	gagcgcaagagagcgctctacgaatactatcatcagatcatcgacaacgagcagagaag	800			
Db	4307	GCACAGACGCCGCTGCGTGGTGGCTTATTGACGAATATGGACGATGGCGCGCAAGTGC	4366			
Oy	801	ccgggaacagcagctatcagaatcgtgtgccaacgycgaggtcaatggtgcagcagatcac	860			
Db	4367	CCAGTGTGCAATTTCAAGACCTTGTGTTTCACGCAAAATCGAGGCGCGTCTTTACA	4426			
Oy	861	agtgacaagacgaagagatggttgagcgtgtacgtgtgctggcgagcgtgatacggltgct	920			
Db	4427	GAGAGCAAGTCCGCGCATTCGTTGCTTTCTTGGTCCGGGCTCGACACGGTTGTC	4486			
Oy	921	aatttcctcagcttcaagcatggaatgcttctgtgcacaaagccggagagatcgccaagagctg	980			
Db	4487	GCCGCATTTGGCTTTGACATAGGCTATCTTGACACGCAATCCAAAGCACCAAGAGTTGTTG	4546			
Oy	981	atcgagcgctcccgagcgatattccagccgctctggaggaactactccgcgcttctcgctg	1040			
Db	4547	CGAAACGAACCGGTGCGGCTTCGGGCTTCGCGTAGAGAACTGCTGGCCCTATTTCACCC	4606			
Oy	1041	gttcgcgaatggcgcaatcccaactcccgatctacgaatttcaatgagctggaactgaaaga	1100			
Db	4607	GTTCCAGATTAATCCCGCTGGGCGCACGAAGACATGAGATTGCAGAGCGCTGCTATCCGTAG	4666			
Oy	1101	ggtgacacagatcctgtctacacgagatgctgtctgcccgtgatatgagcgcaaaacgctgc	1160			
Db	4667	GGTGATTATGTTTCTCTCCCGGATATTTGCTAATCGAGATCCGCTCAATTTAAGTGC	4726			
Oy	1161	ccgatgcacgctgcgaattcagtcgcacaaaggatttcaacacacacacttltggcagcgagc	1220			
Db	4727	CCCAACACTATTCGATTCGGACACAGATAACACACACCCCTTTGGTATGAGACCA	4786			
Oy	1221	catctgtgaccttgcgcacgaacatgcccgcgggaataatcatgltcacctcaagaatg	1280			
Db	4787	CACCTTTGCCAGGAGGCGATCTTCGACGGCGAGAAATCGTCAATGGCCTACGTGAATGG	4846			
Oy	1281	ctgacacaggaattcctgaacttccat	1306			
Db	4847	TTAGCGCGCATCCACGCTTCCGAT	4872			
RESULT	4					
LOCUS	RSPNGR234	34010 bp	DNA	BCT	08-AUG-1996	
DEFINITION	Rhizobium sp. plasmid NGR234a	DNA.				
ACCESSION	268203					
VERSION	268203.1	GI:1486419				
KEYWORDS	BLA; C4-dicarboxylate transporter; CapA; cypA50; fabG; fdxN; fixA; fixB; fixC; fixU; fixX; GapI; glutamate dehydrogenase; lnaA; luxA; nitA; nitB; nitO; Oppp; Oppc; Oppd; Oppf; transposase.					
SOURCE	Rhizobium sp.					
ORGANISM	Rhizobium sp. Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Rhizobiaceae; Rhizobium.					

REFERENCE 1 (bases 1 to 34010)  
AUTHORS Freiberger,C., Perret,X., Broughton,W.J. and Rosenthal,A.  
TITLE Sequencing the 500-kb GC-rich symbiotic replicon of Rhizobium sp.  
NCR234 using dye terminators and a thermostable 'sequenase': a  
beginning  
JOURNAL Genome Res. 6 (7), 590-600 (1996)  
MEDLINE 96389014  
REFERENCE 2 (bases 1 to 34010)  
AUTHORS Freiberger,C.  
TITLE Direct Submission  
JOURNAL Submitted (13-DEC-1995) Christoph Freiberger, Genome Analysis,  
Institute of Molecular, Biotechnology, Beutenbergstrasse 11, Jena,  
07745, Germany  
FEATURES  
source Location/Qualifiers  
1..34010  
/organism="Rhizobium sp."  
/plasmid="pNCR234a"  
/strain="NCR234"  
/db\_xref="taxon:391"  
/clone="cosmid PXB296"  
/clone\_1lb="genomic cosmid library"  
1..625  
/gene="orf1"  
<1..625  
/gene="orf1"  
/note="putative"  
/codon\_start=2  
/transl\_table=11  
/product="OppB homologue"  
/protein\_id="CA92398.1"  
/db\_xref="GI:1486420"  
/db\_xref="SWISS-PROT:Q53191"  
/translation="ILSVTVSGFVGLAARSGGFVDALATFSAIGSVPEVIGYI  
LTFEALIQTRMLPVQGTYSINGCAVPEPEHLILPTVTLISVPTAFIARITRGSMLVTL  
SEDMYRAAKGSGPRAMFPHALKNAGVPIILVIGISPMIGGVLTFTVYNNVEI  
GRLVDAIKRKDPILTIQTVLVLSGLVILNLVDAVTLIDPRIRI"  
628..1503  
/gene="orf2"  
/gene="orf2"  
628..1503  
/gene="orf2"  
/note="putative"  
/codon\_start=1  
/transl\_table=11  
/product="OppC homologue"  
/protein\_id="CA92399.1"  
/db\_xref="GI:1486421"  
/db\_xref="SWISS-PROT:Q53192"  
/translation="MALTTSRASGSPRIQTHVYRIAKRHPLVLDGGTILLILLAL  
AAPLYSGDPLMDPFRLOQPSASMMFGTDNLGRDYFARTYGARISLIYGLLSAVCA  
AVCGLLIGVLAGSRTEFDNIIMRVMDGLMSIPFLLAIALISTGPGIGILYAIAP  
ETPAVTRLVRSVLYSVRSRPYEAALCGARLPRLVRLHILPSTIPLIMQSATVCAS  
AIMEAGLSIFIGVSPSEIPEISMGNMIANSLFLAIALPLTFAPGLCAVTVLVANLIG  
DGLRDMFDPKRRR"  
1505..2512  
/gene="orf3"  
/gene="orf3"  
1505..2512  
/gene="orf3"  
/note="putative: at pos. 1443 also T possible without  
changing the encoded amino acid"  
/codon\_start=1  
/transl\_table=11  
/product="Oppd homologue"  
/protein\_id="CA92400.1"  
/db\_xref="GI:1486422"  
/db\_xref="SWISS-PROT:Q53193"  
/translation="WRMQSNPKADLLLDVADLETHTHYGERSVTALGVNFHVKSSEM  
LGIVSGSGCKSTVALSILRLLPKQTAKTIVGAGVTKGRMLLEISROMEEVGNRI  
FIQDPMTSLSPVHTVROIAEAVOIHGSRSALEKALEMLRLVRIADPERLANN  
PHMSGMRORAIAMALACSPELLIADEPTALDYIQAQIQLRLIVLDKRMGTAVM  
FIYHDGVVAETCORVIMYAGRIIVDAOSTDLFARPTHTYTRALNNSVDRRGSGS  
RIPEIIGVPSLAEPLVGCSEFARCEPFAIGVCEKMPVLSEIQPGHAACWRSNEVYN  
L"  
2509..3570  
gene

CDS  
/gene="orf4"  
2509..3570  
/gene="orf4"  
/note="putative"  
/codon\_start=1  
/transl\_table=11  
/product="Oppf homologue"  
/protein\_id="CA92401.1"  
/db\_xref="GI:1486423"  
/db\_xref="SWISS-PROT:Q53194"  
/translation="MSDALIKVESLTKHPIENGKFTVPPVRAVEDSVFTVOAET  
LCVAGSGCGKSTIARLLMKLLIDPTSGRVLVESTDIALRKNELRAMRRMOWVEDP  
YSSLNRLTAGQIITEPVENFELSLRROQELATDLRKVGMSPEMASSYPSPMSGGQ  
RORIGIARALALPKPSLIADAEVASLDVSOVQOILNLMDLOETGIALIFSHDLAY  
VEHIGHRVAVMYGRIVELSPRDALFVHPPTETALIAAPYPDPARARLEVPLEGE  
VPSVPVPPSGCAFHRPCPLAVDRCRAVEPPLVPMPCGRAVACHVRAPAAAHISPSAL  
ASLALPKTVPGGTILARR"  
3606..4991  
/gene="orf5"  
3606..4991  
/gene="orf5"  
/note="putative"  
/codon\_start=1  
/transl\_table=11  
/product="CapA homologue"  
/protein\_id="CA92402.1"  
/db\_xref="GI:1486424"  
/db\_xref="SWISS-PROT:Q53195"  
/translation="KECSHESGRHSABENDGKYDITGSTATNVYDGFVWAVGVIVS  
RALANGHPGFSEIVELLRAADYTFGNMTFLIFDISFGTPOAEGVAGVHSLPEIG  
PDLKAMGFNIMGRRANNSIDMCVEGKRERSRLIDESGIIHAGVSESAOASARLIET  
ARGRVALLSCATSEFTPMRSRACDPAGCAPAPRGVNALRLRSVYVEPDMESLRKIDA  
LNPQGHKHDREMLVLAGTTRGKGVGYTAAVNRDLADILRNVRGKOYSDFCIT  
NNAHREGMSSEEPADDEQALARKLIDAGDAVYVGHPRPLRGIEIKRRPIYSICNF  
PDDLTETPGADMTVDYDQPOVDYTAETVAAETGTYTAAGTIGALEPEYTESV  
AVSRFEENQIAELRLPIELGYSKRLANGVBSLAPPAISILERLQRLSPFGTRI  
TIEDRGLRL"  
5460..6863  
/gene="orf6"  
5460..6863  
/gene="orf6"  
/note="putative"  
/codon\_start=1  
/transl\_table=11  
/product="BioA homologue"  
/protein\_id="CA92403.1"  
/db\_xref="GI:1486425"  
/db\_xref="SWISS-PROT:Q53196"  
/translation="MTIDIKDISKDRNTVLHPFQOLKDPATGKLEPTIVETGKGR  
IODARNGQLIDGFAGLCYVAVGATKEVAEASISROYRLAYHSYAAHTDELAISD  
RLVKMAPGKMSKVEYGMSSDANETQAKLWYVNNLRGKPTKKIISRERGVHGSVY  
SGSMGMSRHYHDMDLPLQICHTGYPHHYKCANCPETRESARAAALDEMTETIG  
PDNVGAFIAEPVLGTGCTTPPPRGYEAIOAVLKRIIDVLLIDEVITGGRGSMFGS  
OHYGIPEDLITYAKGLTSAVFPPLSASIVGEKVKVLEDDADRVGASHGTYSGHPIG  
AAANAVALDIVERKEDEPGNAREVGYFOALKKEKFNQPLPVEGRVAGLGAIEFVGD  
REKKRRFEDLTKGARVSKAARDRGILIARMBHGDILGAPPLVTYTKKEVDEIVAAAE  
KAVRSYMDLVRDQGL"  
6888..8426  
/gene="orf7"  
6888..8426  
/gene="orf7"  
/note="putative: GNG start codon; alternative start  
codons: 6912; 6927; 7017"  
/codon\_start=1  
/transl\_table=11  
/product="GapD homologue"  
/protein\_id="CA92404.1"  
/db\_xref="GI:1486426"  
/db\_xref="SWISS-PROT:Q53197"  
/translation="MDAGRRRLVINDCMPEGRGRSPLPRRPTRPDPERSGIGINLMA  
TTRAKLMTITSLAPIAVHSPPYDGSLLASVEATPDADIDLATARCAELISNDPRKRR  
ASTLEGAQOVESRDHAEIILIVREAGKTIYQAREVRLCVNTLAKLSAEAAKRNAGEI  
VPFDAYTGEOROGWFTBDPLGIITPAITPYNDPLNLVNAHKLGPALAGNAVMLKPSNL

```
TPESAIKLVLGALREAGLPEBEVITIISHGDELVTYAMTAREVRNVSPTGFGATEAISR
AAGLAKLAMELGNAFVIMNDOPDKAAGCVGAFMAQGCICIAQGLILOSLELFG
RERDAFVAATKRLKAGDPLQEDTDVGPMTLSQAEETEAUVNATIAAGATLILCGE
GSLYHPVLEGTPLTCRLMHEEYFAFVWNLAPEDTLDKIEANNDPDIYSLHACITPND
LVNVALEANRIEYGVWINDSDYRFDAMPFGGFKIGSGKREBSVFRATEDMTPKVVV
INNG"
complement(9781..10860)
/gene="orf8"
complement(9781..10860)
/gene="orf8"
/note="putative: alternative start codons: 10665; 10656"
/codon_start=1
/transl_table=11
/product="transposase homologue"
/protein_id="CAA92405.1"
/db_xref="GI:1486427"
/db_xref="SWISS-PROT:Q53198"
/translation="MGDGRMWSLPEPSRFVSGDSPVPYRAPRGDTCVGLAVIDAP
WGVRAQPVLFRAVPGSAKIAVARMLESGEMTDORHCYAGVDMASESHVFLTDG
RKIGEVFRHGEGLAEMAMLMATGAVGEDIOVAIEVPHGVPEYELLIERGFKVNA
INPKMDREFRDTMAGARDSDDAEVMASALTRDRCRILASPVVYTELAEWSRI
AEDLSAERNRLTRMRERQIMRYFPALLELENDGAWLLDMDYVTPYPPKARIRAT
IAKLKRHRIRRDAPPCARXIAPAAGGRRRDRIRQRPHYAHCPHPPROPAAQTG
ASSAGMPDRQPHRNAGAGAEBA"
1124..12455
/gene="orf9"
1124..12455
/gene="orf9"
/note="putative: alternative start codon: 11220"
/codon_start=1
/transl_table=11
/product="glutamate dehydrogenase homologue"
/protein_id="CAA92406.1"
/db_xref="GI:1486428"
/db_xref="SWISS-PROT:Q53199"
/translation="MSFLDLPEGLPERIIQCNSPYTVREGVRLGRMYSHTGNRSVRE
HCEPVKGIIRYASNDAEEVEALAMLTCSLDVDFPGSKALKIDRGMTPOLE
HTRRPTQEMNKRPPDARAROCVSDIGTGEREMAMDEERANPNPDYSGACVYGL
PLSGKGIAGAEESTGRVOFAIOSLRDTRPELDGRNRNKGSTVYIOGFGNYYAA
RPLESEDDARVYTLERDGYVANPEGLSTELAKKHORIRGSIIGFGANSIADKMGV
EOPCDVLIPAMENAIHAENAEERKAHLVEANGVTEADLILSRGVTIIPDLIV
MAGVYVSFERVKNLTHIPGLMERRRRGRNMTATLAEKMTGESPADMDLEFL
GGAELDLVRSGLDEVMRSTWTRIADLMEQPELDGYRTAAVVASIRQVADL"

Query Match      10 0%; Score 140.8; DB 1; Length 34010;
Best Local Similarity 48.5%; Pred. No. 7.5e-20;
Matches 478; Conservative 0; Mismatches 502; Indels 6; Gaps 3;

QY 324 actgcctgcaagcgagacactgacgcacatcgcgcaactgacatcgctgagcctat 383
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 29617 ACGGGCATGGAGAGGTACTGCTGATTAACAGGGCGAGAGACAGCCCGGCTCTT 29676

QY 384 gaagattacgcgcactttccacgagtgcccgltccatccctcgtagaagcgcgcaagcc 443
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 29677 GAGGATACAGAGACCTTTTCCAGCCATCGCAGACTTTCCTCGCAGCACTTGCGGAACAC 29736

QY 444 taagactcaatccactcgtatgatgcgcccgaagcagcgccaggttcgtgcgtcgcc 503
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 29737 TGGCGGCTCATCCCTCGAATGTGACCCCGGCGCATGAGTGTTCGTGCACTGCTCTA 29796

QY 504 aaccaagtgtgtagcagcggtgtgataaagcttgagagaacgcgataccaggaagctgccc 563
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 29797 AATCCTCTGTTCATCCAGCGGGGTGTGGCGCTGAGCGCACTATACCATGCCAGAGCA 29856

QY 564 tgcctcgtatcgagagcct---gcgcccgcaagagacagtgcaactcaaccgagactaac 620
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 29857 GGCCTGCTCATCCCTCGAATGTGACCAAGAGACACAGCTGTGCTGATGATGAAGATTTC 29916

QY 621 gccgaaccctcccgataccatctcatgctgctgcaagcttaaccgggaagaagatatac 680
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 29917 GCCTTGGCCCTTTACGTTACAGCGTTTCTTCC--CAGCTTCTGTGGAGACTTTCACAGGCGATC 29975

QY 681 ccggaactgaaatacctaacgatacgaaccgctccggaatgcaagcatgacatgacctcgca 740
```

```

Db 29976 CGAAGTACTTGTGGCTGGGTAAAGCCATTTGCTCCAGCGCAACGCAAAAGCAAG-- 30033
QY 741 gaagccaagaagcgctctcagactatctataccgatactcgaaacgaagcagaag 800
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 30034 GCAGCAACCCCGCTGGCTGGTGGCTTTATTGACCAAAATGCGAGGAAGCCGCCCAACTCG 30093

QY 801 ccggaaccgagcgtacagcatcgatctgccaagcgccaggttcaatggcgaccgatacc 860
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 30094 CCACTGTCAATTTCAAGACCTTGTGTTCAAGCGCAAAATCGAGGGCCCTCTTAACA 30153

QY 861 agtgaagcaagcaagagatgctgctgttactgltcgcgccctgatacgtgctc 920
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 30154 GAAGAGAGAAATCCGGCGGCAATCGGTGCTTTTCTTGTGTCGGGGGCTCGACACGGTTGCC 30213

QY 921 aatttcagcttcagatgagatctcgcgcaaaagcccggaacatcgccaggaagctg 980
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 30214 GCCGCCATTGGCTTTGACATGGCGCTTATCTTGACCGCAATCCAAAGCACACAGAGTTCTTG 30273

QY 981 atcgagcgctccgagcgatltccagcgcgctcgaggaaactactccgscgtctctgctg 1040
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 30274 CGAAACGAAACCGGCTCGGCTCGGCTGCGCTGAGGAACCTGCTGCGGCTATTTCACC 30333

QY 1041 gtgacgcatgscgcacatccatccatccatgacagttcatgagcgtgcaactgaaaga 1100
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 30334 GTTCAGATTAATCCGCGTGGCGGCAAGAGACATGAGTTCGAAGCGCTCCCTATCCGTGAG 30393

QY 1101 ggtgaacagatcctgctacgcagatgctgctgagcgtgagagcgcaaaacgctgc 1160
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 30394 GGTGATTAATGTTCTCTGCTCCGCGGATGATTCGATTCGATGATCCGCTGTAATGAATGC 30453

QY 1161 ccgattgacgtgactcagtcagtcgcaaaagtttcaaacacacacttggcagcgagc 1220
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 30454 CCCAACCTATCGATCTGGACAGCAGAGATTAACCAAGCACCGGCTTGGCTATGAGCCA 30513

QY 1221 catctgctcgttcgcaagcactgcccgcgcggaataatcatcgtcaaccctcaagaaatg 1280
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 30514 CACCTTTGCCACGAGAGCATTCGCGCAGGCGAGAAATCGATTCGCTACAGGATTCG 30573

QY 1281 ctgaaccaggaatcctgacttccat 1306
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 30574 TTAGCGCGCATCCACGCTTCCGCAAT 30599

RESULT 5
LOCUS A79351 236165 bp DNA PAT 20-OCT-1999
DEFINITION Sequence 2 from Patent EP0818465.
ACCESSION A79351
VERSION A79351.1 GI:6092394
KEYWORDS
SOURCE .
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 236165)
AUTHORS Broughton,W.J. and Perret,X.P.
TITLE GENOMIC SEQUENCE OF RHIZOBIUM SP. NGR234 SYMBIOTIC PLASMID
JOURNAL Patent: EP 0818465-A 14-JAN-1998;
BIOLOG MOLECULAIRE DES PLANTES (CH); INST OF MOLECULAR BIOTECHNOLOG
(DE)
FEATURES
source 1..236165
location/Qualifiers
BASE COUNT 48909 a 68589 c 69313 g 49354 t
ORIGIN
Query Match      10 0%; Score 140.8; DB 5; Length 236165;
Best Local Similarity 48.5%; Pred. No. 6.2e-20;
Matches 478; Conservative 0; Mismatches 502; Indels 6; Gaps 3;

QY 324 actgcctgcaagcgagacactgacgcacatcgcgcaactgacatcgctgagcctat 383
```

[illegible]

VERSION	A93003.1	GI:6741431
KEYWORDS	unidentified.	
SOURCE	unidentified.	
ORGANISM	unidentified.	
REFERENCE	1 (bases 1 to 236165)	
AUTHORS	Broughton, W.J. and Perret, X.P.	
TITLE	GENOMIC SEQUENCE OF RHIZOBIUM SP. NGR 234 SYMBIOTIC PLASMID	
JOURNAL	Patent: WO 9802560-A 22-JAN-1998; BIOLOG MOLECULAIRE DES PLANTES (CH); BROUGHTON WILLIAM JOHN (CH)	
FEATURES	Location/Qualifiers	
SOURCE	1..236165	
	/organism="unidentified"	
	/db_xref="taxon:32644"	
BASE COUNT	48909 a 68589 c 69313 g 49354 t	
ORIGIN		
Query Match	10.0%;	Score 140.8; DB 5; Length 236165;
Best Local Similarity	48.5%;	Pred. No. 6.2e-20;
Matches	478; Conservative	0; Mismatches 502; Indels 6; Gaps 3
QY	324	actgcgtcgaacggcggaacatcgtgatcgcacatcgcgcgcgaactatccgttaagcctat 383
Db	146785	ACGGCCGATGACGAGGATCACTTCCTGGGATATACACGGCGCAGACCGCCGGGCTCTT 146844
QY	384	gaagattacgcacatttcacgcagatgcccgtlcatcctcgttgaagccgcgaagcc 443
Db	146845	GAGGATACGAAACCTTTTCCAGCCATCGCAGATCTTGCTCGGACATTGGCGAACC 146904
QY	444	taagacttcatcccaactcgaatggatccgcgcgaagcgccagtttcgttcgtcgtgcc 503
Db	146905	TGCGCGGTCATCCCTCGAATTGGACCCCGCCCATGGAGATGTTTGTGCATCTGTA 146964
QY	504	aaccaagtggttggcatgcgcgttggttgataagcttgggaaccggtatccagaagctcgccc 563
Db	146965	AATCCCTGTTCTCAATCCAGCGGGGCGTGGCGCTCGACCCGACCATATCATCTCCACAGCA 147024
QY	564	tgcctgcgtgatcgaagacct--gcgccgcgaagaacagtgcaacttccagaggaactac 620
Db	147025	GGCGGGCTGATGACTGCTTTCGGAAGGAAGACCACTGTGACGTATGAAGATTTTC 147084
QY	621	gccgaacccctccgatacgaatcttcatgctgctcgcaggttcaaccggaagaagatalc 680
Db	147085	GCCTTGCCCTTACGTCGCTTTTCCCT--CAGCTTTGGGAGACTTTCCAGAGGGGATTC 147143
QY	681	ccgcacttgaatlaactaaacgatacgaatgagaccgttcgcgatggcagcatgaccttcgca 740
Db	147144	CGAAGTACTTGTGGCTGGGTAAAGGAGTTTGTCCACGGCAACGAGAAACGAAG-- 147201
QY	741	gaggcgaagaagcgctcctcaagacatctgataccgatalccgaacaacgagcgagaag 800
Db	147202	GCAGACGCCCGCTCGGTCTGTGGCTTTATATGACGAATATGCGATGCGCGCAGATGG 147261
QY	801	ccggaacccgaacgtatcaagcatcgtctgcgaacgycgcaggttcaatggtgcacgcatcacc 860
Db	147262	CCAGCTGTGATTCATGACCTTGTGCTGCTTCAAGGGGAATAATGAGAGGCCGCTCTTAACA 147321
QY	861	agtgaacgaacgaagaagatgtgtgacctgttactggttcgcgcgccttgatacgtgtgtc 920
Db	147322	GAAAGGAAAGTCCGCGGATCGGTGTGCTTTCTTGGTGCGGGGCTCGACACGGTTGGCC 147381
QY	921	aatttcctcagcttaagatgagttccctcgcgcgaacaaagcccggaacatcgcaggaagctg 980
Db	147382	GCCGCCATTGGCTTTCACATGGCTTATCTTGACGCATATCAACACACACAGAGTTGTTG 147441
QY	981	atcgagcgctccgaagcgtatctccagccgcttgaggagaactactcgcgcgtcttcgcgtg 1040
Db	147442	CGAAACGAACCGGCTCGGCTCGGCGCTCGCGCTGAGAACTCTCTCGCGCCTATTTCAMCC 147501
QY	1041	gttcgcgatctggccgcatctccactccgattacgaagtttccatgycgttgaaatgaagaa 1100
Db	147502	GTTTCGATTAATCTCGGCTGGCCGGAAGGACATCGAGTTTCGAAGGGCGTGGCTTATCCGTGG 147561

OY	1101	ggtgcacggatccctcgtactaccgcagatgctgtcgttcgtacctgatltga/cgcgaaacgcctcyc	1160
Db	147562	GSTGATTTATGTTCCTCGTCCGGCAGTAGATTGCTAATCGAATACCGTCTGAATTTAAAGTGC	147621
OY	1161	cggatgcgaagctcgacttaagtgcgccaaaagtttcaacacacactcttggtgcacgcgcgc	1220
Db	147622	CCCAACTATTCGATCTTGCGACAGACAGGATTAACAACGACACCCTTGCGCATGACCA	147681
OY	1221	catctgtccttgcgcgaacactggtcccgcgcggaatacatcgtcacctccaagaattg	1280
Db	147682	CACCTTTGCCACGAGGAGCGCATCTGCCACGGCGAGAANAATGTCATTTGGCTACTGTAATGG	147741
OY	1281	ctgaccaggattcctgacttccat	1306
Db	147742	TTAGCGCGCATCCACGCTTTCGCGAT	147767
RESULT	7		
SERCP450A			
LOCUS			
DEFINITION	SERCP450A	3078 bp DNA	BCT 26-APR-1993
ACCESSION	Saccharopolyspora erythraea ORF 1 gene, partial cds; cytochrome P-450 gene, complete cds; ORF 2 gene, partial cds.		
VERSION	M83110		
KEYWORDS	M83110.1 GI:152682		
SOURCE	cytochrome P450.		
ORGANISM	Saccharopolyspora erythraea (library: NRRL 2338) DNA. Saccharopolyspora erythraea Bacteria; Firmicutes; Actinobacteria; Actinomycetales; Actinomycetales; Pseudonocardineae; Pseudonocardiaceae; Saccharopolyspora.		
REFERENCE	1 (bases 1 to 3078)		
AUTHORS	Anderesen,J.F. and Hutchinson,C.R.		
TITLE	Characterization of Saccharopolyspora erythraea cytochrome P-450 genes and enzymes, including 6-deoxyerythronolide B hydroxylase J. Bacteriol. 174, 725-735 (1992)		
JOURNAL	92121109		
MEDLINE			
FEATURES			
source	location/qualifiers		
	1..3078		
	/organism="Saccharopolyspora erythraea"		
	/db_xref="taxon:1836"		
	/tissue_1fb="NRRL 2338"		
	<1..859		
	/note="ORF homologous to Escherichia coli and Salmonella typhimurium dnaE gene; putative"		
	typhimurium dnaE gene; putative"		
	/codon_start=2		
	/transl_table=1		
	/protein_id="AAA26482.1"		
	/db_xref="GI:152683"		
	/translation="AGGGRGHRQAOPAGARVPVRRRGRRRRGVVVAAGAPVHVG RVAAQDALLIABGADGLRIGPSRGSRRAARAARPGRHARRARRARGCGPCGDPR DDLDPPAPHODERPYPVDROPDGRRRGCAVLEPVLIDVRRLLPARGHRODEPHOR AKGHQHLKRRGAGHGUGKDRPNOPRVADGAGSGGVNARQAHLAAGLHGIG PGAEEAAGAARGHPRAVRYLRLLRVDERAARGAARGVALRDGAASRGAASPLRA PRPRLRGs"		
	949..2166		
	/codon_start=1		
	/transl_table=1		
	/product="cytochrome P-450"		
	/protein_id="AAA26483.1"		
	/db_xref="GI:152684"		
	/translation="MTTGEVVDLIAFPDAFAODRNHYARKREPYVORIPTVNGLDAM LTITVEDYKALIDLPRIAKDGPRTQIIERYLADERRPGSPDLGPHMINTDPDHFT RKRLUYKAFTARAREEGIRPIETIDTDLLDRLAGREVDLIIDFAFLPTVISLM GVEBSRDDERSWTNVLVDSQPAAQASVAMEYLTELAKRTPEGDDLITALLE AVEGBGRDISBELIAMVELLVAGHEFTVNLIGCVSLTELRNPDLAARNDSILALLE AIETEELRESYVANGTERHTPAEARFEDVVPECELWVALGAANDGERFEEDPRFD ITRETGTVAAGHGIFHCVGAAALARLEAQIAVGRLLRPFDLRMAASPDLLRMRFSTVL MRGLEXLIPVRGA"		
	complement(2365..>3078)		
	/note="ORF 2"		
	/transl_table=1		
	/codon_start=1		
	/transl_table=1		

[illegible]

Db 2077 CGGATGGCCGCGGAGACCTGCGGCGGTTACAGCGTC 2121

RESULT 8  
AF147703/c  
LOCUS  
DEFINITION  
AF147703 5833 bp DNA BCT 01-JUL-1999  
Streptomyces fradiae demethylmacrocin O-methyltransferase Ty1E  
(Ty1E), hypothetical NDP-hexose 4-ketoreductase Ty1D (Ty1D),  
tylodoxin (Ty1H), cytochrome P450 Ty1HI (Ty1HI),  
macrocin-O-methyltransferase Ty1F (Ty1F), and putative NDP-hexose  
3-epimerase Ty1J (Ty1J) genes, complete cds.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
SOURCE

AF147703 1 GI:5305783  
Streptomyces fradiae.  
Bacteria: Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
1 (bases 1 to 5833)  
Bate,N.,. and Cundliffe,E.  
The mycinose-biosynthetic genes of Streptomyces fradiae, producer  
of tylosin  
Unpublished  
2 (bases 1 to 5833)  
Bate,N.,. and Cundliffe,E.  
Direct Submission  
Submitted (30-APR-1999) Biochemistry, University of Leicester,  
University Road, Leicester LE1 7RH, UK  
Location/Qualifiers  
1..5833  
/organism="Streptomyces fradiae"  
/strain="J59235"  
/db\_xref="taxon:1906"  
complement(72..1259)  
/gene="ty1E"  
/note="orf24\*"  
complement(72..1259)  
/gene="ty1E"  
/function="converts demethylmacrocin to macrocin during  
tylosin production"  
/note="2'-'-O-methyltransferase"  
/codon\_start=1  
/transl\_table=11  
/product="demethylmacrocin O-methyltransferase Ty1E"  
/protein\_id="A041815.1"  
/db\_xref="GI:5305784"  
/translation="MAVOKKATLVKQIRAAAGAAADVREIVAEHGEPAVAVYDEI  
VSRAPRVADVIVELAVRSGDALVPRRLAQAQAPVRRAPDDGCVAMRYEELD  
ELVRELFGPCREKAAAGTGTTPVATSGHIDYFLAAQATATVLAGCTSAKDL  
NELTSHYLPKMGSLHMFPHYDRHREYNEVRVLEIGIGQHPHMGGSILRMK  
HFEHRLIYGLDIEDKSHAEORITTVGQNDPQGLTTELARYGPEDIYIDGSHIN  
EHVRSFHALFPHVRPGRGLYIEDLMTAWPQGGSDPGKSDLSLIGKLSVDSLO  
HOLPEPDSGRSPCYADRHVYGLHINLAIFIEGVNSEGICPGMITRDRDALVAASG  
GA"  
complement(1277..2287)  
/gene="ty1D"  
/note="orf23\*"  
complement(1277..2287)  
/gene="ty1D"  
/function="involved in synthesis of NDP-deoxyallase during  
tylosin production"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical NDP-hexose 4-ketoreductase Ty1D"  
/protein\_id="A041816.1"  
/db\_xref="GI:5305785"  
/translation="MRAAGDWAGRTVVVTGALGFIGSHYAEKIALGARYVGLHNSAR  
PAAGAPAAPAAGRSCLVADLCDEAATRAQEMAGADVTHCAGLDONAQYKNDH  
SASVLDANVGTAVNTARDTGAAGAVLLSTEVYCAPRDSAPAREDEIRRYPHAG  
NGYVLSKICEIMAEHLGAEEGSRIFRVGNGVYGRDGGGRTRTVPSMVAARA  
ETFEIWDGROFRTFHVHEDIVATLRMDADKVPYPMNIGSPREYSILDIARTVSAV  
GKGGITRTYTPDKSPPAORLDSNRKDVTCGEPRLRVGIEOTITIRKOHHRPSAP  
HS"

gene complement(2284..2529)  
/gene="ty1HI"  
/note="orf22\*"  
complement(2284..2529)  
/gene="ty1HI"  
/note="ferredoxin; Ty1HI"  
/codon\_start=1  
/transl\_table=11  
/product="tylodoxin"  
/protein\_id="A041817.1"  
/db\_xref="GI:5305786"  
/translation="MRVRIIDTGCRCVAGAGCCERAPPTVFNODEDVGCVLDKRPAPVW  
EEVREAEDLCPPARAVALSCDGTGAGAGAAAPPTGRDA"  
complement(2565..3875)  
/gene="ty1HI"  
/note="orf21\*"  
complement(2565..3875)  
/gene="ty1HI"  
/function="hydroxylation of polyketide ring at C-23 during  
tylosin production"  
/codon\_start=1  
/transl\_table=11  
/product="cytochrome P450 Ty1HI"  
/protein\_id="A041818.1"  
/db\_xref="GI:5305787"  
/translation="MAMADPTVFSLRPQHVSSGDPARPSOKGILLPAARANDTDEA  
GRSITAMPVARTCPSPPEQVAAALRAEPIARAELMDGAPVWLISRODHVRLADPR  
VSHHPAKLRLSPSDCEAASASRLTLPDPDHGALRGHPIPERGLRVELRPSYQI  
VTGLDDLTARGDADLADLAFAPMAQYICRLDIPYEDKDTFQERTQATRPAAE  
EALFALLELDVYLDRLISGKTGREGSDGMGVAVARCGGLSHADVLNAYLLAAG  
HETTSAMVMSVYLVLIQHPAMRELTVNGLLPGAVDELLRYSIADGLRSATADIE  
IGHTTIRAGDGLVFLAANRDEAVFSEEAQDIHRSARRHYAFGYGPHQCIGQNLAR  
MELEVALGAVLEELPALRPTTQVAGLRKSDSAVFVEYELPVAM"  
4090..4860  
/gene="ty1F"  
/note="orf20\*"  
4090..4860  
/gene="ty1F"  
/function="converts macrocin to tylosin"  
/note="3'-'-O-methyltransferase"  
/codon\_start=1  
/transl\_table=11  
/product="macrocin O-methyltransferase Ty1F"  
/protein\_id="A041819.1"  
/db\_xref="GI:5305788"  
/translation="MAPSPDHARDLYIELKKVSVNVIEDPTHVAGMTTDSFQDTS  
RESGEYPRVAHMTGLKRLDNLHRLAVVEDGVGDIETGVWRGACIPARGLIN  
ATGQADRTVWADSFQFPELTGSDHPLVEIDLHOYNEAVDLPSTSETVRENFARG  
LDDNVRFIAGWFKDTMPAPVAKQLAVMRLDGDSYGATMDVLDLSLYERLSPGYVTD  
DVCIIPACREAVHDFRDLGIRDTIHRIDQAGVWRHSG"  
4897..5514  
/gene="ty1J"  
/note="orf19\*"  
4897..5514  
/gene="ty1J"  
/function="synthesis of NDP-deoxyallase during tylosin  
production"  
/codon\_start=1  
/transl\_table=11  
/product="putative NDP-hexose 3-epimerase Ty1J"  
/protein\_id="A041820.1"  
/db\_xref="GI:5305789"  
/translation="MRDTRPLGIEGAWVIOPEIHPDRGEEFHAMFOSSEPRRLTNGS  
FSYPOVNIAVSRKGLRGTHFSEVPPGQAKYSACVAGAVEVYVDRLGSPYQWRA  
VPDEYNRTAVVYPACLGAFVALDTRTTLVYLCSESYAAREHAYNPDPDLATAMP  
EDVELLSRDRAPTLAEARKGILPSTQETREHGLPEHARR"

BASE COUNT 799 a 2072 c 2098 g 864 t  
ORIGIN

Query Match 7.9%; Score 110.6; DB 2; Length 5833;  
Best Local Similarity 47.4%; Pred. No. 1,8e-13;  
Matches 397; Conservative 0; Mismatches 434; Indels 6; Gaps 2;

[illegible]

RESULT	9	
AF055922/c		
LOCUS		
DEFINITION	AF055922 12905 bp DNA BCT 04-MAY-1999 Streptomyces fradiae dipeptidyl carboxypeptidase (ddca), tyrosin resistance protein (tlrB), glyoxylaloxanase (tylA), tyrosin methyltransferase (tylE), 4-ketoreductase (tylD), ferredoxin (tylH2), cytochrome P450 (tylH1), macrocin O-methyltransferase (tylF), epimerase (tylJ), acyl-CoA oxidase (tylP), and butyrolactone receptor (tylQ) genes, complete cds.	
ACCESSION	AF055922	
VERSION		
KEYWORDS	AF055922.1 GI:4210925	
SOURCE		
ORGANISM	Streptomyces fradiae. Streptomyces fradiae Bacteria: Firmicutes: Actinobacteria: Actinobacteridae:	

REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES
1	Alcalá-Medina, J.	Streptomyces	Streptomyces	Streptomyces	1	Alcalá-Medina, J.	Streptomyces	Streptomyces	Streptomyces
2	(bases 1 to 12905)				2	(bases 1 to 12905)			
3	Fouces, R., Mellado, E., Díez, B. and Barredo, J.L.	The tylosin biosynthetic cluster from Streptomyces fradiae: genetic organization of the left region	Microbiology	145 (Pt 4), 855-868 (1999)	3	Fouces, R., Mellado, E., Díez, B. and Barredo, J.L.	Direct Submission	Submitted (27-Mar-1998)	Ingenteria Genetica, Antibioticos S.A., Avda. Antibioticos 59-61, Leon 24080, Spain
4	Location/Qualifiers				4	Location/Qualifiers			

gene complement(235. .1431)  
/gene="ddca"  
CDS complement(235. .1431)  
/gene="ddca"  
/note="similar to Bacillus cereus alkaline D-peptidase and to Streptomyces serine dipeptidase"  
/codon\_start=1  
/transl\_table=11  
/product="dipeptidyl carboxypeptidase"  
/protein\_id="ABD12161.1"  
/db\_xref="GI:4210926"  
/translation="MRKPLAILMAVTAVIGYQAAAPKTAGKGTAEAGHRPTAAI  
QREADALGAPVGLADLRPGRGVKVRSGGPDIDACTPYPMKARAPFISGLTTPVPA  
ATVLDLVEGGKSLSDIDTDHMLPGIVSRGNDGSAITGMLRHRSGLRDVPPLPYTY  
FLEKEGPRRRRTVPRQOAVRLALHPPVPAPRTGWSSTNTYITLAGITVKATGHSK  
QHEKVRITVPIGLRHTIAPDTPFTIPGVHAGTORFEKGLPADPAPRSGEVDVAT  
VNPSSVSGAAGEMIISTTDEGNRFLALLGGELALPAELAEKKTVVRAGEPDSAMPVVR  
YGLGMAANSCGVWFHGGDIPGFKTRNGVTCGDSRSVSVNTQSVPEPVARRP  
GRRIP"  
gene 1788. .2630  
/gene="tlrB"  
CDS 1788. .2630  
/note="tlrB"  
/note="similar to Micromonospora griseorubida Myra  
protein"  
/codon\_start=1  
/transl\_table=11  
/product="tylosin resistance protein"  
/protein\_id="ABD12162.1"  
/db\_xref="GI:4210927"  
/translation="MRKNVRYLRCPHCAPLRRSDRTLRCENGHTEDVAROGYVNL  
RLPRTAAATDTMVAARAALDSCGHVAPLTERLACTACRAGAGAPDCVVAIGGTTG  
HARVLEEDAEEDGLLMSKPVARRAARLPRASAAVDVADPTLPLRGGESPAALG  
FAPNDPELTRILRPGCTLLVYTPQODHLAEVDAIGLRPRDKRKGSLAQQLAPHEE  
ANGOERLRTLLDHALGRVAVAMPSSMHODPELARRIAELPEIHEVTLSTVFIVC  
RLP"  
gene complement(2700. .3968)  
/gene="tylN"  
CDS complement(2700. .3968)  
/gene="tylN"  
/note="thiamosyl transferase"  
/codon\_start=1  
/transl\_table=11  
/product="glycosyltransferase"  
/protein\_id="ABD12163.1"  
/db\_xref="GI:4210928"  
/translation="MRALITWGSRGDVPQPVVALGCTLRANGHEVYLGAEPALPLVE  
QAGLERATPGDGFETWPEVVEVTLRGRPMARDLMLKALPAPEYDQEVLDRLERAG  
ECVADLVNAPLVTVTALGEPTSPMLVSWMNPNTSTWTFPAVESQGRMGPLTPLYNRLL  
THMRVLEDEMGWRARAEVDFRGRGLIPFGKSSPLRRJGHRPHLLYFPSPVLPK  
MGQCGHATVYEMPDQGWRRSPLEDFLADGEPVLLTIGSTVPHNGQEEVEVAANA  
ANGARBRLLVYGGPEGALRGALRPSVDSADYSLMDRTPAAVYVHGFGGTTADAVAGCP  
OYLVYVPADHPNMAKRLRKGTAARPVPLAAMNRBALAASVRYATVDPAMVARRRLG  
EVAVAERGENCAVLEEMAEKRTTAHPG"  
gene complement(4014. .6229)  
/gene="tylE"  
CDS complement(4014. .5201)  
/gene="tylE"





Oy	883	gtgagcctgttactgtgtcggcgccctcgatagtggtgtaatttctcaagcttcagatg	942
Db	7012	cggctctcctgcttgccgccccgggacacgagaccacgacgacgctggtacagatgacgttc	6953
Oy	943	agctctgaccaaagcccgagacatgcacgaagctatctgagctcccgagctatc	1002
Db	6952	tgcgttctctacacagaccctccacggccttgccgggaactcaccttgatattccggcctgttc	6893
Oy	1003	cagcgacttgcgaagaaactactcgcgcgttcttcgctgttgacgatg---gcgcatacc	1059
Db	6892	cggcgccgctggagacaacatgctggccctatgtctgattgcgcacggcgctgcgcctcgg	6833
Oy	1060	tcaactccgattagagattcaatgctgcgtgcaactyaaagaaggtatccagatcctgtac	1119
Db	6832	ccaccggcgacatcgcacgtcgcacgccacaccatccggccggcgacacggcctgtgcttcc	6773
Oy	1120	cgcagatgtgtctgcgtcgctgagatgagcgcgaanaagcctgcgcgcgattgcagctgactta	1179
Db	6772	tgcctggccggcccaaccggagcgaagggccgttttctccgaaccgagggcctttcgacattcc	6713
Oy	1180	gtcgcacaaaggtttcacacacacacttltgcgcacgcagccactatgtcttgccttgcagc	1239
Db	6712	accggtccggcgaaggcgcatgtgccttgcgttacggaaccacacagctgcctgggacaga	6653
Oy	1240	acctgcccgcgcgggaatcatcgtcacccctcaagaaagatgctgcacaggaattcctg	1296
Db	6652	acctgcgccgagtgacattggaatgagggcctctggcgccgtgtgagagggctgcggc	6596

RESULT	10
AFI27374	
LOCUS	AFI27374
DEFINITION	AP127374 63734 bp DNA BCT 27-MAY-1999 Streptomyces lavendulae Lina homolog, cytochrome P450 hydroxylase ORF4, cytochrome P450 hydroxylase OFB3, MltT (mltT), Mts (mts), Mtr (mtr), Mito (mito), Mtp (mtp), Mto (mto), Mitn (mitn), Mtm (mtm), Mtlc (mtlc), Micp (micp), Mtk (mtk), Mti (mti), Mth (mth), MigC (migC), Mif (mif), Mite (mite), Mld (mld), MitC (mitC), MIB (mIB), MlC (mlC), MlA (mlA), MmCA (mmCA), MmCB (mmCB), MmCC (mmCC), MmCD (mmCD), MmCE (mmCE), MmCF (mmCF), MmCG (mmCG), MMCH (mmCH), MMCI (mmCI), MmcU (mmcU), MmCK (mmCK), MmCL (mmCL), MmCM (mmCM), MmCN (mmCN), MmcJ (mmCJ), MmCX (mmCX), MmCV (mmCV), MmCR (mmCR), MmCS (mmCS), MmCT (mmCT), MmCU (mmCU)
ACCESSION	AF127374 complete cds; and unknown genes.
VERSION	AF127374
KEYWORDS	GI:4731328
SOURCE	Streptomyces lavendulae. Streptomyces lavendulae.
ORGANISM	Bacteria; Filicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE	1 (bases 1 to 63734) Mao,Y., Varoglu,M. and Sherman,D.H. Molecular characterization and analysis of the biosynthetic gene cluster for the antitumor antibiotic mitomycin C from Streptomyces lavendulae NRRL 2564 <i>Javendulae</i> NRRL 2564 <i>J Chem. Biol.</i> 6 (4), 251-263 (1999)
AUTHORS	2 (bases 1 to 63734) Mao,Y.Q., Varoglu,M. and Sherman,D.H.
TITLE	Direct Submission
JOURNAL	Submitted (10-FEB-1999) Microbiology, Uni. of Minnesota, Box196, 420 Delaware St. SE., Minneapolis, MN 55455, USA 3 (bases 1 to 63734) Mao,Y.Q., Varoglu,M. and Sherman,D.H.
REFERENCE	Direct Submission
AUTHORS	Submitted (27-MAY-1999) Microbiology, Uni. of Minnesota, Box196, 420 Delaware St. SE., Minneapolis, MN 55455, USA
TITLE	Amino acid sequence updated by submitter
JOURNAL	Location/Qualifiers
FEATURES	1..63734
REMARK	/organism="Streptomyces lavendulae"
SOURCE	

CDS

/strain="NRRL\_2564"

/db\_xref="caxon:1914"

123. .1367

/function="dehydrogenase"

/note="ORF6"

/codon\_start=1

/transl\_table=11

/product="unknown"

/protein\_id="AAD28446.1"

/db\_xref="GI:4731329"

/translation="MRNTLFEIDGIDIPICGVPEOMASVIRREYCGPHQAFSEYVD  
VPKVGQCALVLVMAAGINYNKNWASLQGPVDVISAQOKGSHSDHIGSGESGVV  
AVGSEVTVAAAGDEVILSGCOMTEFLAADIRLGADPMWSSQSWGEGNGSPAQAL  
VDDYQKARGKPGRLTWEENACPLILGATAYRQLCGMPHDVDPDPVYLWGAGAGLCSM  
AIOITRARGGIPVAVVSEDERARVCRELGAQGTINRDFPHMGRPLDIGHENMGMT  
EYVRAFGREMEVLEGRSPRIVLHESQOATIPISWILCNAGVYVTCGTTGYNDI  
DRELFMOKRKLOGSHFANLRQCRDVIHMVANQGLDPCLSMTGGFDDIGKANLMDND  
OHQGNQAVLVNAPRTGLTTPA"

1428. .1958

/note="ORF5"

/codon\_start=1

/transl\_table=11

/product="Lina homology"

/protein\_id="AAD28447.1"

/db\_xref="GI:4731330"

/translation="MDKLDILMSEREIRAVLQRYCGLDRLDELYKSAAYHEDANDR  
GVIRNANDFPYQIYPLRLDAVYGTGLTHTLHGSTLEIDGADYESVCTAAYHRESGI  
KRVEQFACRYVDFEFRDGVKIRARLVLDNDFSLAGVPLDPAEAQAGFNRSHDLTD  
ASYOVLPRGPDAPTL"

2006. .3229

/codon\_start=1

/transl\_table=11

/product="cytochrome P450 hydroxylase ORF4"

/protein\_id="AAD28448.1"

/db\_xref="GI:4731331"

/translation="MSDTQEHAPTLPRQRCPSPEPPELBEKRTDPISNMRADSP  
GWLTLRHADVRALADPGVSSHGPKAPQPMWRNLAPEMRACHYLPGLIFMDPDHRY  
RLTLKRLTMAIRKLEPRLQIVETLDAEAGQVDLVQSPALPIPLVLICELHG  
IRYEREEFMDVLRALMDATPEELALGRMEFPMKLAAKRAVPGDLSLILAD  
PDDAPADLTDEIGAGIVGLM.IAGHETSANMLGVTGTLLENDDOMALNDODISLD  
AVELLRHQITVQOGLPGCVTQMEINGAHQVKTGESLTLPLANDPARNVFPDPDID  
ITREINPLAHGHGHIHLCGLKMLARVENRQAMRWLVTRFGLRMAAPEDIRWRDQI  
YGVGNLVPTWDEAK"

3226. .4458

/codon\_start=1

/transl\_table=11

/product="cytochrome P450 hydroxylase ORF3"

/protein\_id="AAD28449.1"

/db\_xref="GI:4731332"

/translation="MTCGEAAVRCPCGACGAPAYPEHAPDRLEPDPYMERLRERPL  
ORVTLPYGCEAMLATRYODRAVAFADRRFSROLAVAGPGRPLPHOPPPAVLSVGP  
DHARLRRLVGVGFTPRPREDMRPLIORTAGSLDAMEEMKRPADIVDEDSLPAVSMI  
CELLGVPPEDKRRFCWMSDALVLTTHATPAQVORYKQMDYVJGGLVAKQRRVPLND  
IGSLVTADEEDKLTGEBLLALAEALILAGETSASOIPAEVLVLFPHQDLRIAND  
HDILPDAAVEELRLRPVIGTVGFPRTAEVLELGGVAVRAGEVLRVMSGCAANDPELF  
TDPDLDLARRPNPLGCGAPRHCLCAQLARVELQITLTTLPRRYPRILRAVBEESL  
SMKEELMWRGHTMPVTW"

4552. .5658

/function="esterase"

/note="ORF2"

/codon\_start=1

/transl\_table=11

/product="unknown"

/protein\_id="AAD28450.1"

/db\_xref="GI:4731333"

/translation="MSTIDEWENSTKESAGDPAALRRLTDVVARGCAALCYMRGCT  
VLWDSFECSSDSLFLVYAATKPAALALVALAERGLIGDRPAVEYVMPFAHGSGD  
VTVRIVLDRAGVAPVGVGIVRTKRTADWGESVADLQSPKKGGEVVAVHMSGCV  
ITGELVQVTVRSPRDRVYTSLEFAPLGINLHNGLPSASAPRRVYPARAAHPSMPNOM  
MSNRGTYQAVITYPSAGLSGTRQAQMARYQMLMBGSGLDGIRLVLPETVEARKPSNDG  
GIDNSLKRVPWMSHGFMJGGGPPRGVLSVLTGTSPPSAFGHAGTTSSVYADPTRE  
LVIALTSINIOPEFAGIERLRREVSDIALGACEAG"

5992. .6849

CDS

CDS

```

/strain="NRRL 2564"
/db_xref="taxon:1914"
123. .1367
/function="dehydrogenase"
/note="ORF6"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAD28446.1"
/db_xref="GI:47313329"
/translation="MNTLTPELGDIPDICTVPEBPVYSIAKRVQSGVSGEDPHICGSBSGVV
VAKVPGSALQILVYMAAGTIVNYMASLQGPVDYISAKRQVSGVSGEDPHICGSBSGVV
AVGEGTQVAVGDEVLTISCCQMTETADIRLGLADPMTSSQSWGEGVNGSPAFAL
VDYQCHNPRLGTWEACFLTGTATAVROLGOMPQHVPRGDPVLINGAGAGLSG
ALQIDIRARQGIPIVAVYVSDERARVCELSAQGTINLDPDHMRPLDIDHEMGMT
EVSARARREWEELGSRPRVLTIESGQATPTSYLCDNMGMYTIGCGTYNDI
DLRLPMRQKRLQSGSHRPLALROCRDVIHVAANOQLDPCLSMTGCFPDICRANLMDN
QHPQGNQAVLYVAPRGLTTPA"
1428. .1958
/note="ORF5"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAD28447.1"
/db_xref="GI:47313330"
/translation="MOKLIDLWSEEREIRAVLYQRYCHGLRDLDEELVKSAYHEPDADR
GVIRGAHNPVQVQIVLDELDAATGLTHTLHGSTIEIDGAAGVESCTAVYHESQDI
KVEQVAGVYVVEERESQGVKIRLVINDPSLAGEVLPDPAEQAGFNPSHRLDLD
ASYQVPLNGPAPLT"
2006. .3229
/codon_start=1
/transl_table=11
/product="cytochrome P450 hydroxylase ORF4"
/protein_id="AAD28448.1"
/db_xref="GI:47313331"
/translation="MSDTQGHAPLTFRQRTCPSPPELLELRDPISRMKPADSP
GNULTIRHADVRAALADPGVSSHGKAPQPMRNLAIPMAREHYLPGLIFMDPPDHRY
RLRLTKRPMKRIKLEPRILEQVETLDAENQGTVDLYOSFALPIPLVLCIELH
IYEEKEEPMDFMRLQALDAVPEELGALGARNNEPMKLAARKRANPGDGLSHLAG
DDPADRLDTELAGVGLMGLTGHSTSNMGLGVGYTTLLENDDQALLRDQISLDR
AVIELLRHQTIVQGLPGRCTVDMELIAGHOVGTGSESLASPRANDPAVFPDPRLD
ITREHNPILAFSGHILHLCMLGLARVEMKQMGVLVTRPGLMAAPEDIRKDDQI
VGVVYVNLPTWMEAK"
3226. .4458
/codon_start=1
/transl_table=11
/product="cytochrome P450 hydroxylase ORF3"
/protein_id="AAD28449.1"
/db_xref="GI:47313332"
/translation="MGCPAAVAGCCPGAGAPAYPFHAPDLREPPVEYLERERPL
GVNLTPEGGEAMLATPEVOVRAVADRRFSROLAVPFAPRFLPHQPPDPAVLVSQG
DHARLRIVYGVKTPRPRVDDMRPLQRTADGLLDANEBGPRADLVEDSSIPRAYSMT
CELLVGPDPDKRFFCVSSDALITTAHTTAQVRSADIMQMHQDILGGLVARORVAPL
ISGLVVARQEBKLELGEGLVRLAEALLAGVYFSDIYFIVLVRHQLERIAND
HQLIDPAVELLRLFVPIGVNDPPRATDEVELGAVLVAGETVYVSKAANDPELF
DPDELDLARRNPRLGFGAGPHNCLGAQLARVELITLTTLPRRVRLRAVBEESL
SKKEGLMVGCMTPMTVM"
4552. .5658
/function="esterase"
/note="ORF2"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAD28450.1"
/db_xref="GI:47313333"
/translation="MSTIDWEHSTKEACMDPAALRLRLDVVARGSAQOLCYMRGCT
VIVDSFGSSSSSLFVYATKPAVALAYNALAERLILDPDRVATYVQAFARHSGC
VIVRHLVAGSAPVQRCIGVPMRTAGDMSRVLDEQSRPKMPGGEVVAHFMSGF
ITGELVRYGTGVSFSDVYSELFAPLGLDLMNGLGSWMPRHVRAARAHPSMPPQW
MNSNRTROAVYLPSCGLSTQAQMAARFTQMLDEGSLDGLIVLRPTTVEARPSND
GIDASLIRKRVVNSHAGLGGPDPGRGLSNVLTGRYSVDPAFGHAGCTTSVWADPRE
VLAATLSNQPEFGAGIERTLRVSDIALGACACAG"
5992. .6849

```

5992. .6849

/function="transcriptional activator"  
 /note="ORF1"  
 /codon\_start=1  
 /transl\_table=11  
 /product="unknown"  
 /protein\_id="AAD28451.1"  
 /db\_xref="GI:4731334"  
 /translation="MLNLPKMERAHHPSPQVGLPEVRSAGGAGTGANVSGIV  
 RLLAALTAIRGQAMSTERILKEVWADNPAPTIDKRAVAVALRLRVLGDNCEWMLLT  
 RSSGVALIDIPDHLDAVRAETLVREGRAALACDPVAARARILTRALDQNEPEVADANR  
 AISTYSQTELEIANRSEAVOAHIDARLELGHOBELVGELEARNPLHEPHLQJAM  
 LALYSGSQAEELAAVMQLRQALAMENLGIIDRGRLQELHLRLILAADGDLTGSFAP  
 AEPILVRS"  
 complement(6854. .7666)  
 /gene="mitr"  
 /complement(6854. .7666)  
 /gene="mitr"  
 /function="aminoquinolate dehydrogenase"  
 /codon\_start=1  
 /transl\_table=11  
 /product="MitR"  
 /protein\_id="AAD28452.1"  
 /db\_xref="GI:4731335"  
 /translation="MEFIFPAAVGSAGATRLAYVLDGPVQYKAPGILLPLISGSLDA  
 VVYPRVHVARDLAEYVEGLKRIGINDGLITVYPRKALCGIADLGGRAALLIGTANAM  
 RKEPDEPVAENFEDLGVQGIQAAGHTVRORHVALVAGAGSASITATLMDAAVSV  
 EVHDDRQALDALLRLGSRRRDGIRALGPDLLAOPVAVNATPLGMRSDPLPFDP  
 RRPRAVVVDVYMKRHEFTALLSAATAGRAVHHGITHLEQVPCYRAAFGWP"  
 7895. .8842  
 /gene="mits"  
 7895. .8842  
 /gene="mits"  
 /function="kinase"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Mits"  
 /protein\_id="AAD28453.1"  
 /db\_xref="GI:4731336"  
 /translation="MTGDTDGAAGGGDVTFRMPAAGDVTADILLARVKGILLGHREDP  
 LAGVAVPAMICIDAGVTRTPMGRPSMAGLMLTAFGQLPGTFVACADDLALAEAE  
 SAAACRHLILVYGVGTGIGGGIVHGSRAMPGRGSCGVHVVVDNRSGPRDCCRACG  
 VOAVVSGPATLRRAAERRGRETGEDELASGARLHAPMAAVDESAALATAVNGIOE  
 LAHPVLVYGGFPAAGVPGYVASAHAHYERTTRGTDPVRYRPAVILGSRSLHALLL  
 AEAHGKGRPPRESVSDVSVFEGVTDRAVGRSD"  
 8916. .10460  
 /gene="mitr"  
 8916. .10460  
 /gene="mitr"  
 /note="similar to mitomycin C oxidase MctA"  
 /codon\_start=1  
 /transl\_table=11  
 /product="MitR"  
 /protein\_id="AAD28454.1"  
 /db\_xref="GI:4731337"  
 /translation="MLDRRSVIRVGAGAAAANAAGFATGATAAAGVLPGRGAASG  
 VMAULRRHLSSELVPADGTGERAKRLYSQDFGIRPQAVATRTEDVKTTLAEFQ  
 DHALVITRSGSGHSPGYSTTDGIVLDVSGFHAVSLRNTVVMGAGTQOVDALTALEP  
 RSVAAASNCAGVCPGCFVOCGGLGMOGRKFGMACDRLSVARVLADGRAVTAATSP  
 PLFMAAGCGGGNGFVVTGPELRPTDVSIVSNLTMPMSARVILEAMOHMIIIDP  
 RLGAAMVOPDAGCTGPRVYVVGAMTGADALTPVLDLSVAVSGASQATRSKAKLS  
 OHDMAAOYGCAADLTPEOCHTVGYSPEALFRQNFSDMRNRLFSSRAIGQGGVEITL  
 EAADRAQOFLRSLFALGGAANRPDRTTVAVHRDEEYLGFSIGLNDPEYTAEDERR  
 LGRAAVACRFLTIDPHSNGESYONFIDPELDWMKSAYVAENYVRLAAVKAAYDPHRLF  
 SPQAV"  
 10607. .11101  
 /gene="mitQ"  
 10607. .11101  
 /gene="mitQ"  
 /function="putative regulator"  
 /codon\_start=1  
 /transl\_table=11  
 /product="MitQ"  
 /protein\_id="AAD28455.1"

gene	/db_xref="GI:4731338"	
CDS	/translation="MERELTRMPVESERRERCRDQVMRLIVLEGAEPALCYDPEE DWVAIPVSTDDLRARVEALRLRGAASRPEDVNGVLRFRMRALLSPEARIVARL AESYAEEVYARDDLRPPGRTPVSRNALDLHMIRRRRLAALGLRVTYVGRGVLESE AEGV"	
	11157. .12188	/gene="m1p"
	11157. .12188	/gene="m1p"
	/function="aminodHQ synthase"	
	/codon_start=1	
	/transl_table=11	
	/product="M1p"	
Query Match	7.4%;	Score 104.2; DB 2; Length 63734;
Best Local Similarity	47.4%;	Pred. No. 3,1e-12;
Matches	452;	Conservative 0; Mismatches 483; Indels 18; Gaps 4;
QY	452	cattccacactcogatgatcgccgcgcgagacagccagttcgtgcgtcgtcccaaccaagt 511
DB	2272	CTTCCCTGATCTTCATGAGACCCGCCCGACACACCCTGACCCGCCCTGCTCACCAGATG 2331
QY	512	ggttgcgcgtccgcggtggtgtgaagcttgagagaccgagatccagagagcgtcgtctct 571
DB	2332	GTTCACCATGCGGGGCATCTCCGAACTCGAACCAGATCGAGATCGTACCGGAAC 2391
QY	572	gatcgagagcctctgcgcccgcaagagacgtgcacttccaccgagaga---ctacgcgcgaacc 628
DB	2392	CCTCGACCGCATGAGAGGCCACGGCGCGGACCGTGCACCTGGTGCGAGTCTTCGCCCTGCC 2451
QY	629	cttcccgatacgatcttcaatgatcgtcgtcgcgaagttacgcgaagaagatalccgcacct 688
DB	2452	GATCCCGCTGCTGTCATCTGCGAAGCTGATGGCATCCGCTACGAGGAGCGCGAGAGATT 2511
QY	689	gaataacctaaagatcagatgaccgcgtccgagatgca---gcatagactctcgagagagc 745
DB	2512	CATGCACATGCTCTCCGCACTCAAGCCCTTGAGCCCAACGCCGAGAACTGGGGCCCT 2571
QY	746	caagagagcgctctcagatactatgataccgatactcagagcaacgacagacagagccgag 805
DB	2572	CGGCGCCAGGATGAAAGAGATTTCATGATGAGCTGCCCGCGCCAGAGCGCGGAAACCCGG 2631
QY	806	aacgcagcgtatcagatcgtctgccaacgagccaggttcaatggcgagccgatccaccagtga 865
DB	2632	CGACGACCTGCTCAACCCACCTCCGCCACGACCCCGACCCCGCGGCTACCGAGACT 2691
QY	866	cgaagccgaagagatgctgtgaccttcaactgtgtggcgagccgtgagatgagtgatatt 925
DB	2692	GGAGATCCGCCGCGATCGCGTCTGATCTCATTCGCCGGGACCGAGACTCGGGCCAAACT 2751
QY	926	cctcagcttcagcatgtagtctcgtgccaaaagcccgagacatcgacagagagctgatcga 985
DB	2752	GCTGGGCGTGGGACCTTACACCTCTGTGGAGAACCGACAGTGGGCCGTGCTCCGGA 2811
QY	986	gcgtcccgagcgtatctcaagcgcgtcttgagagaactactccgcgtcttcgtctgttc 1045
DB	2812	CGACATCAGCCCTGATCGACCGGGCCGTGGAGAGCTGCTGGCCACACGACACTGTGTCA 2871
QY	1046	cgatggc-----gcatactcactccgcgattacgagtctcatgtgcgtgcgacatgaaagaa 1099
DB	2872	GCAGGAGCCTCGCCCGCGGTACCCCGGAGACTGGAGATCGCCGGGACCAAGGTAAAGC 2931
QY	1100	aggctgacagatcctgcgtacacgcagaatgctgtcgtcgtctgatatgagcgcgaanaacgctg 1159
DB	2932	CGGGGAGTCCCTCTGGCTCGCTCCGCCGCCGCCCAACCGCGACCCCGCGTCTTCCCGGA 2991
QY	1160	cccgaatgacgttcgaacttcagtcgccaanaagtttccacacacacactttggccacgagcg 1219
DB	2992	CCCCAGCGCCCTCGACATACACCGGAGAGACAAACCCGACCTCGGCTTCCGGACAGGAT 3051
QY	1220	ccactgtgctcttgccagcagcactctgcccgcgcggaataatcgtcactccctcaagaatg 1279
DB	3052	CCACTCTGCTCTGGCCATGAGAGACTCGCCCGGGTGGAGATGGCCACAGGCTGGCGGCGCT 3111

QY	1280	gcgcacagagatctctgacttctccattccattcgtcccgaggttcccgatctcagacaagacgcg	1339
DB	3112	CGCTACAGCGCTCTCCCGCGCTCGCATGGCCGCCGCCGACAGACATCCGCTGGCGCCGA	3171
QY	1340	C-----acgtcaccagcgcgtcgaagcaccctccctcgatctgcatccgcgcga	1386
DB	3172	CGACCAAGATCGCTCTACGGCGCTGTACAACTCCCGCGTGACTCGGACGACGAGCCCA	3224
RESULT	11		
LOCUS	SPU08223		
DEFINITION	SPU08223 10167 bp DNA BCT 04-FEB-2000		
ACCESSION	Streptomyces fradiae cytochrome P-450, dNTP-glucose synthase, dNTP-glucose dehydratase, thioesterase, TyLCVI (tylCVI), and TYLR (tylR) genes, complete cds; and unknown gene.		
VERSION	U08223 AF145042		
KEYWORDS	U08223.2 GI:6849140		
SOURCE	Streptomyces fradiae.		
ORGANISM	Streptomyces fradiae		
REFERENCE	Streptomyces fradiae		
AUTHORS	Bacteria: Filumicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomyces; Streptomyces; Streptomyces		
TITLE	1 (bases 1 to 10167)		
JOURNAL	Merson-Davies, L.A. and Cundliffe, E.		
MEDLINE	Analysis of five tylosin biosynthetic genes from the tylIIA region of the Streptomyces fradiae genome		
REFERENCE	Mol. Microbiol. 13, 349-355 (1994)		
AUTHORS	2 (bases 7082 to 10167)		
TITLE	Bate, N., Butler, A.R., Gandecha, A.R. and Cundliffe, E.		
JOURNAL	Multiple regulatory genes in the tylosin biosynthetic cluster of Streptomyces fradiae		
MEDLINE	Chem. Biol. 6 (9), 617-624 (1999)		
REFERENCE	3 (bases 7082 to 10167)		
AUTHORS	Bate, N., Butler, A.R., Smith, I.P. and Cundliffe, E.		
TITLE	The mycarose-biosynthetic genes of Streptomyces fradiae, producer of tylosin		
JOURNAL	Microbiology 146 (Pt 1), 139-146 (2000)		
MEDLINE	20121747		
REFERENCE	4 (bases 1 to 10167)		
AUTHORS	Merson-Davies, L.A.		
TITLE	Direct Submission		
JOURNAL	Submitted (31-MAR-1994) Louise A. Merson-Davies, Biochemistry, University of Leicester, Adrian Bldg., University Road, Leicester, LE1 7RH, UK		
REFERENCE	5 (bases 1 to 10167)		
AUTHORS	Butler, A.R.		
TITLE	Direct Submission		
JOURNAL	Submitted (23-AUG-1999) Biochemistry, University of Leicester, Adrian Bldg., University Road, Leicester, LE1 7RH, UK		
REMARK	Sequence update by submitter		
COMMENT	On Feb 1, 2000 this sequence version replaced gi:473596.		
FEATURES	Location/Qualifiers		
SOURCE	1..10167		
	/organism="Streptomyces fradiae"		
	/strain="759235"		
	/db_xref="taxon:1906"		
	944..2197		
	/note="ORF1"		
	/codon_start=1		
	/transl_table=11		
	/product="cytochrome P-450"		
	/protein_id="AAA21341.1"		
	/db_xref="GI:473597"		
	/translation="MTQTLEAEKPEPEERTANSVHPLAOPAGAGAGLLEWFAAR		
	EAPVWDSRQAMQVRYDDYLTSTNPLFSSDSPAYPEELAILMGPGFGGI		
	PPRHPKRLKLSOATPRRIATLEPRIEITGGLDREKQIDVSLDAPLPIYI		
	IAELGIPAEKBDLEKWDVLTINMGEMVPLPDDFSMTKPAIKEMQDYLRIAI		
	KRPETPDLMGSLIAEVEGRLLDDELYIVALLTACHISATILGNLFLVLDH		
	EQSALRDRLITGAIETLIRNSPFNNITFLLEDDIDIGHPMKAGOMVAWIAS		
	NRQAHFSDPTDFVRQPNKMSGHHIHLGSLFLAEKVPLELTFDESDVR		

CDS  
 EHDEVEFEYEDDELTAARRLPVTYTHN"  
 2677..3843  
 /note="ORF2, GTG start codon"  
 /codon\_start=1  
 /transl\_table=11  
 /product="unknown"  
 /protein\_id="AAA21342.1"  
 /db\_xref="GI:473598"  
 /transl\_table="WTGLPRPAVRVPEHDLRDVHAATGVSEIGGALLRYAARGRYLL  
 GALLAFAEEREAECGNAMHCVAVSGSDDDALALMALVGEGBGVIPSHTFASMLA  
 VSATGATVPVPEGDPPEPGALFLDPDRLEALTPRTBAVRVHLHYGPVLDLDPG  
 AFAEHPGALVADAQATARYRGRGRIGSGHTAASPGRNIGALGGGAVVSDPEL  
 ADRLRLNTEGAREKRYRHEERGTNSRLDELOAALSTKYLPLDAMNRRREITANRGE  
 ALAGCPGTVYEGRAEPRVWHQIVYLRSTFRDLRLRLAEGVETLVAHVAAAGAY  
 AGAGCCPAGGPGPRAERLAGEVLSLPDGLPDEAVEVYIAAVSALADSMEGP"  
 3840..4751  
 /note="ORF3: GTG start codon"  
 /codon\_start=1  
 /transl\_table=11  
 /product="GTPD-glucose synthase"  
 /protein\_id="AAA21343.1"  
 /db\_xref="GI:473599"  
 /transl\_table="MNDRPRAMKGIILAGSGSTRRLPLNGTISKOLLPYVDPMIY  
 PLSVIMAGIREIOIISKDLPLRSLLEGDBDLGISTAYDRPSPYIAEPFLIGA  
 RHIGDDAALLIGDNPVPHGPGFPSSYLCTGVARLDGCEFGYPADAHRYVGEIDSG  
 RLILSEEPMPRLPEGRHRLVLYTNDVLEORTITSPARELEITVDNKKYLLBGRRA  
 HGAGVAVMLPGLMEGTCHDSLLAGOVOLLEIOPRGERIACIEIEMRMGFIABECCYRIG  
 OELRSSYSGYIIDVAMRGAADSRAQ"  
 4812..5813  
 /note="ORF4"  
 /codon\_start=1  
 /transl\_table=11  
 /product="GTPD-glucose dehydratase"  
 /protein\_id="AAA21344.1"  
 /db\_xref="GI:473600"  
 /transl\_table="NRVLVTCGAGFICSHPTGOLLTCATYVDLCATRTVYLDKLTACN  
 PANEHVAQHDPLEVRGDIADHGMNRMIMVGLVVFPAESHVDRISSSEAFRT  
 NEGVRLYLQAAVADGAVGRVPHIISTDEVYSGIAGSMPEDHPAPNPSPYATNAASDL  
 LVALHRTYGLDVRTYRCSSNNYGRROYEKAUPLFTTNLDGLPVPYGGGWTREU  
 HHDDICRVALVAGAGRGVYINIGGTELELNAELTDRILELGCADRSALRAVADRQ  
 HDRRSVDTTKITRELGVAAPRTGITELAGTVANRDNRMAMWELTKSPGRELERA"  
 5875..6636  
 /note="ORF5, GTG start codon"  
 /codon\_start=1  
 /transl\_table=11  
 /product="thioesterase"  
 /protein\_id="AAA21345.1"  
 /db\_xref="GI:473601"  
 /transl\_table="WMLRCYAPVPCTPAHRLVCPPHAGGSARAYRPALELAAAGVET  
 HAVQYPGSDRRKEPFAFTLELAEERVLPELRLILADPDVPAVLEFGSHGAAVADET  
 ALLHRSQAPRPAAGLILSGRAPADARETLEHLLGDEBLLAEIRLGGTSGALADEE  
 VLRWLPLIRDDVYAVAGRVRRVPGPRGCCLTVTVTGADPNVTLPEAMRELTLTGAF  
 ALRVPGGHFPLNDREAVNCRITIEETLRHGSKAH"  
 6763..8130  
 /gene="tylCVI"  
 6763..8130  
 /gene="tylCVI"  
 /function="required for mycarose biosynthesis"  
 /note="dNTP hexose 2,3-dehydratase: tyl orf6"  
 /codon\_start=1  
 /transl\_table=11  
 /product="tylCVI"  
 /protein\_id="AAF29379.1"  
 /db\_xref="GI:6849141"  
 /transl\_table="HTLPGDITRMARBRAGRGADVALLPEPAMRGWHFEDPNLRHE  
 SGRFPSITGLNVRKGDGPGVWGPILVQPEVGLGIVAREIHGVFLHFLVQAMBEGN  
 INTLQISTVYATRSNFGVHARGRIGRFLDLTEPGARVLYLVLSQEMQVETLARN  
 RNMYVEIDPEDELTEGDEFPWLTLGQIRALLFTLANLNMADARSYLACILPAGGGOVAG  
 DDFAVLNRSTSDSAPTLGSAEELSGMFGSVYALREELVORLIPLAVANAHGHEHGD  
 ELAHSNGHHPVLAADITASRREYARRNSQPLIOGRHPSLLALTYKRVDGYLHLYVAR  
 VDVGHLNLAELAPVYHCHRPADHRAPEAPAPYLDVVLASPAQFRDYTVQSEEGRF  
 YHARRKVAIVEPESFDAASGTDEDTAGDYAMWTFQGLTELHLAHGNVYNELRTLAC  
 AHATY"  
 gene  
 CDS

Query Match 7.3%; Score 103; DB 2; Length 10167;  
Best Local Similarity 47.2%; Pred. No. 6.7e-12;  
Matches 414; Conservative 0; Mismatches 430; Indels 33; Gaps 2;

QY	1155	gctgagccgcatgacac---gtcagcttcagltgcgcaaaaggtlltcacacaccacccttggc	1211
Db	1964	ATGAAGCGCCGGTGCAGATGCTGCTGCATGATGCCGTCGCCCAATTCCGACAGCGCCAC	1963
QY	1212	cacggcagccatctgttccttgcgcgcaacctgtgcccgcgggaatcatcattccctc	1271
Db	2024	CACGGCACCACACACACGCTGGGCTCGTTCCTGGCCGGCGGCGGACCAAGTGTTCCTC	2083
QY	1272	aaggaatgctgaccagatctcctgacttccattg	1308
Db	2084	GAGCTGTCTTCGACGAGATTCTCCGATTACCGCGCTCG	2120
RESULT	12		
SSU65940			
LOCUS	SSU65940	2500 bp	DNA
DEFINITION	Streptomycetes sp. cyt P-450-9-deoxo-FK506 hydroxylase (fkbd) and 31-O-demethyl-FK506 methyltransferase (fkdm) genes, complete cds.	BCT	05-SEP-1996
ACCESSION	U65940		
VERSION	U65940.1		GI:1522692
KEYWORDS			
SOURCE	Streptomycetes sp.		
ORGANISM	Streptomycetes sp.		
REFERENCE	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomycetes. 1 (bases 1 to 2500)		
AUTHORS	Motamedi,H., Shafilee,A., Cal,S.J., Streicher,S.L., Arison,B.H. and Miller,R.R.		
TITLE	Characterization of methyltransferase and hydroxylase genes involved in the biosynthesis of the immunosuppressants FK506 and FK520		
JOURNAL	J. Bacteriol.	178 (17),	5243-5248 (1996)
MEDLINE	96359380		
REFERENCE	2 (bases 1 to 2500)		
AUTHORS	Motamedi,H.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-AUG-1996)		
FEATURES	Natural Products Drug Discovery, Merck Labs, P.O. Box 2000 Bldg R80Y-225, Rahway, NJ 07065, USA		
SOURCE	Location/Qualifiers		
	1..2500		
	/organism="Streptomycetes sp."		
	/strain="W46858"		
	/db_xref="taxon:1931"		
	/note="FK506 gene cluster"		
	196..1362		
	/gene="fkbd"		
	196..1362		
	/gene="fkdm"		
	/note="cytochrome P-450 hydroxylase"		
	/codon_start=1		
	/transl_table=11		
	/product="cyt P-450-9-deoxo-FK506 hydroxylase e"		
	/protein_id="A244359.1"		
	/db_xref="GI:1522693"		
	/translation="MSFDPACGTPOAGRCFPAIODAHRAITDSATGSPDLFGIKHMLVAAAEKIVTNDPRESSAAPSMLPDRRPGWSGMDVPEHSPYRQKINGDFTLRARROSEFVFEADACLDEAAGPGADLVPGTAKRLPSLVINALYGLTPREGAVLETRARGITGSDUDSVKTLTDFEFAHLELVRAKRDGDDLHRLAAGEGEIPLSDDEATGVFETLLFAGHDSVOQWGYCYLLALSLHPDEQELRTRPILIDAVSEMLRFLPNVONGVPRCAVADVLDRGVRSVGDNIPLVSTARDEVPDPPTFVSAPTEGNAFGHGVHPCPGVAVRLRLIKVACLRLERFPVRLYGVDPVNMNGIGLSPALRLRTWGAA"		
	1359..2141		
	/gene="fkdm"		
	1359..2141		
	/gene="fkdm"		
	/codon_start=1		
	/transl_table=11		
	/product="31-O-demethyl-FK506 methyltransferase"		
	/protein_id="A244360.1"		
	/db_xref="GI:1522694"		
	/translation="MSDVYETLLPNCATVAHNAGEAOFLYREIFTDSDCYLRHGVEL		



QY	935	cagccttgagatcttcctgcccacaaagcccgagagatctgcgaagatcgtacgcagcgtctccga	994
Db	777	CAGCCTCATCACCGTCTCTGGAACACCCCGACGACGACGCGCGCCGCGCCGACCGCCGTC	836
QY	995	gcgatlctcaagccgctctgaggaactac-----tccggcgctctcgtctglttgcga	1048
Db	837	CCTCGTCCCGGAGGGGCTGAGGAACCTGCGGCTCTCGGCGCATCCGGGAGCATCCGGG	896
QY	1049	tggccgcatctcactcctccgatlacgagtttcaatgagcgtgcaactgaagaagtgcaca	1108
Db	897	CGCGCGCATCCGCGCACCGCGGACATCGAGATCGACGACGACGATCCATCCGAGCGGTAAGG	956
QY	1109	galtctgctaacgcagatgctgctgctctgtagagcgcggaacgcgtcgcgatgca	1168
Db	957	ACTGATTCGTACCAACTCCATCTCCATCCGACCGGACAGCTTGGCTTGAGAAACCCGACCG	1016
QY	1169	cgctcgaactcaagtcgcacaaaggattcaacacacacacacttggccagcgacgcatctg	1228
Db	1017	CCTCGATGTGACACCGCTGCGGACACGCGCACACACCTCTCTCTCGGGGTACGGGGTCCACAGTG	1076
QY	1229	ccttgccagcaactctgcccgcgcgggaataatcgtctcaactcgaagaatgctgacga	1288
Db	1077	CCTGGGACCAAGAACCTGGCGCCGCTCTGCAACTCGAATCATCTCTACACGTTGTTCGACCG	1136
QY	1289	galtcctgactctccatctgc	1310
Db	1137	CATTCCGACCTTCGCGCTGCGC	1158
RESULT	14		
LOCUS	STMCYP450		
DEFINITION	STMCYP450 1274 bp DNA BCT 07-FEB-1999		
ACCESSION	STPctomyces carthophilus gene for cytochrome P-450sca-2, complete cds.		
VERSION	D30815		
KEYWORDS	D30815.1 GI:1072316		
SOURCE	Cytochrome P-450sca-2.		
ORGANISM	Streptomyces carthophilus (Isolate:SANK 62585) DNA.		
REFERENCE	Streptomyces carthophilus		
AUTHORS	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.		
TITLE	1 (bases 1 to 1274)		
JOURNAL	Serizawa, N.		
REFERENCE	Submitted (25-MAY-1994) to the DDBJ/EMBL/Genbank databases.		
AUTHORS	NOBUFUSA Serizawa, SANRYO Co., Ltd., Biomedical Research		
TITLE	Labortories; Hiromachi 1-2-36, Shimagawa, Tokyo 140, Japan		
JOURNAL	(Tel.:03-3492-3131(ex.3330), Fax:03-5436-8565)		
REFERENCE	2 (bases 1 to 1274)		
AUTHORS	Watanabe, I., Nara, F. and Serizawa, N.		
TITLE	Cloning, characterization and expression of the gene encoding		
JOURNAL	cytochrome P-450sca-2 from Streptomyces carthophilus involved in		
MEDLINE	production of pravastatin, a specific HMG-CoA reductase inhibitor		
FEATURES	Gene 163 (1), 81-85 (1995)		
source	96001248		
	Location/Qualifiers		
	1..1274		
	/organism="Streptomyces carthophilus"		
	/isolate="SANK 62585"		
	/db_xref="taxon:44059"		
	35..1267		
	/codon_start=1		
	/transl_table=1		
	/product="cytochrome P-450sca-2"		
	/protein_id="BA06492.1"		
	/db_xref="GI:1072317"		
	/translation="MTETKATTFELTSSOENAPFADRTCPYQLPTAYSRLEKDEPDL		
	RPVLYLGGRRAMVYTKHKAARLLADRLSDRIHAFPAISPEFKFQSGPAFIGM		
	DEPHGRBRMTISEFTVKRIKGMRPVERVHEFTIDDMIACGTALVGFALPVS		
	MYICNMGVPAADHEPRQDASKRLVQAVVDASAAARDDEPRYIDGLITLESPTG		
	LLGLVTHQIADGETIDRAELISTRLALLVAGHETVAMTSLSTVITLLEHPQQRALKA		
	DPSTVPAQVEELLKRVLAIDADGRIATVADIIDGQLIRAGEGIVTNSIANRDSVFE		

Query Match	7.0%	Score 97.6;	DB 1;	Length 1274;
Best Local Similarity	46.3%	Pred. No. 1.1e-10;		
Matches 399;	Conservative 0;	Mismatches 454;	Indels 9;	Gaps 2;

  

Query	Match	Best Local Similarity	Matches 399;	Conservative 0;	Mismatches 454;	Indels 9;	Gaps 2;
458	caaccgcgatgtgtcccgccagcagccgacggtttctgtgcgttgcg	46.3%	0;	0;	0;	0;	0;
331	CATCGGGAATGATATCCCCGAGACAGGGAGCGGTGCGCGCATGACGATCAGCAGATTAC	46.3%	0;	0;	0;	0;	0;
515	tgagcatgcgcggttgtgtatgaagcttggagaacccggtatccagaagctgtgcctgtcat	46.3%	0;	0;	0;	0;	0;
391	CGTGAAGCGCATCAAGGGCATATCGCCCGAGCTCGAAGCATCGGTCAGCGGCTTCATCGA	46.3%	0;	0;	0;	0;	0;
575	cgagagcccttgcgccccgaagaacagttgcaacttccacgaggaactaacgcccccttccc	46.3%	0;	0;	0;	0;	0;
451	CGACATGCTCGCGGGGAGACCCACCGCGATCTGGTCAAGCATTCGCGCCGCTGCTCC	46.3%	0;	0;	0;	0;	0;
635	gatacgcatcttcatcgtcgtctcgcaggtctaacggaagaatatccccaacttgaata	46.3%	0;	0;	0;	0;	0;
511	GTTCATATGGTGAATCTGCATATCTCTGGGCTCCCTCAGCCGACACAGATCTTCCAGGA	46.3%	0;	0;	0;	0;	0;
695	ccctaacggaatcagatgaaccggtctcgcagatgtgcaagcaatgaacttgcagagccaaagagc	46.3%	0;	0;	0;	0;	0;
571	CGCGAGCAAGCGCCCTGGTGGACAGCGGATGGAGCCGACAGTGCCTGCTGCGCCGGAGCA	46.3%	0;	0;	0;	0;	0;
755	gctctaggaactatcttgataccgatatgagacgaacgcagagcagaagccggaaaccgacgc	46.3%	0;	0;	0;	0;	0;
631	CTTTCAGAGCGCTACCTGAGACGGGCTGATACCAAGCTTGAGTCCGAACCCGGGAGCGGCT	46.3%	0;	0;	0;	0;	0;
815	tatcgaagactcgtttgtcgaacgagcaggtcaatggagagcaactaacagttgacagccaa	46.3%	0;	0;	0;	0;	0;
691	CCTCGGCAATCTGTGTACACCCACACAGCTGGCGGACGGCAGATCCAGCCGCGGAGCTAT	46.3%	0;	0;	0;	0;	0;
875	gaagatctgttgcgttacttgctgcgcgcgtctgatacgttggatcaatttctcgaact	46.3%	0;	0;	0;	0;	0;
751	CTCCACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	46.3%	0;	0;	0;	0;	0;
935	cagcatgaagttcctcgtgcgcaaaagccggagacatgcagaagactgtatcgagctccga	46.3%	0;	0;	0;	0;	0;
811	CAGCGTCAATCAACCTGCTCGAAGACCCCGACAGCACCCGCTGCGCGCGCGGACCGCTC	46.3%	0;	0;	0;	0;	0;
995	gagtatctcaagccggttgcgaggaactac-----tcggcgacttctgcggtttgcgca	46.3%	0;	0;	0;	0;	0;
871	CTCTGATGCCGAGGCGGTGCGAGAACTCTGCGGCTCTTGACCATCGGAGCATCGCGGG	46.3%	0;	0;	0;	0;	0;
1049	tgagcgcatactcccaactccgataaagatttcaatgtgcgtgcaacttgaagaagttgacca	46.3%	0;	0;	0;	0;	0;
931	CGGCGGCAATCGGACCGGCGGACATCGAGATGAGCGGAGACGTCAATCGCGCGCGGAGG	46.3%	0;	0;	0;	0;	0;
1109	gactctgtacccgagatgtgtgtccttgcgtatgtagcgcaaaagccttgcgcgatgtca	46.3%	0;	0;	0;	0;	0;
991	AGTGAATCTGTCACCACTCATCGCCAAACCGGAGACGTTGCTGTTTCGAGAACCCGAGACG	46.3%	0;	0;	0;	0;	0;
1169	cgctgacttaagttgcgaaaagtttccacacacaccccttggccagcgacgcatctgtg	46.3%	0;	0;	0;	0;	0;
1051	CTCTGATGTGACCGCTCGGACGACGCCACCATCTCTCTTCCTGGGATGAGGATGACAGTG	46.3%	0;	0;	0;	0;	0;
1229	cccttgcagcagcctgcgcgcgcgggaatcatcgtcacccctcaagaagatgtgcagcag	46.3%	0;	0;	0;	0;	0;
1111	CTGTGGGCGCAAGACCTGCGCGCGCTGCGAAGCTGGAAGTCACTCTCACCGGTGTTCGACCG	46.3%	0;	0;	0;	0;	0;
1289	gattctgacttctcaltgcc	46.3%	0;	0;	0;	0;	0;
1171	CATTCGACCGCTGCGCGCTGCGC	46.3%	0;	0;	0;	0;	0;

  

RESULT	15
PSECAMABA	

LOCUS PSECAMABA 1950 bp DNA BCT 01-FEB-2000  
DEFINITION Pseudomonas putida gene for cytochrome P-450cam, NADH-putidaredoxin  
reductase, putidaredoxin, partial cds and complete cds.  
ACCESSION D00528  
VERSION D00528.1 GI:216870  
KEYWORDS putidaredoxin; camC; NADH-putidaredoxin reductase; camA; camB;  
camphor; cytochrome P-450cam.  
SOURCE Pseudomonas putida (strain:ATCC17453 (strain PpG1)) DNA,  
clone:puj45.  
ORGANISM Pseudomonas putida  
Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;  
Pseudomonas.  
REFERENCE 1 (bases 1 to 1950)  
AUTHORS Koga, H., Yamaguchi, E., Matsunaga, K., Aramaki, H. and Horiuchi, T.  
TITLE Cloning and nucleotide sequences of NADH-putidaredoxin reductase  
gene (camA) and putidaredoxin gene (camB) involved in cytochrome  
P-450cam hydroxylase of Pseudomonas putida  
J. Biochem. 106 (5), 831-836 (1989)  
JOURNAL 90130389  
MEDLINE  
COMMENT Submitted in computer readable form by H. Koga on 13-Jan-1990.  
FEATURES  
source  
1..1950  
location/Qualifiers  
/organism="Pseudomonas putida"  
/strain="ATCC17453 (strain PpG1)"  
/db\_xref="taxon:303"  
/clone="puj45"  
/note="2 bp upstream of SmaI site.; total cellular DNA"  
1..93  
/gene="camC"  
<1..93  
/gene="camC"  
/gene="camC"  
/note="carboxy terminus"  
/codon\_start=1  
/transl\_table=11  
/product="cytochrome P-450cam"  
/protein\_id="BAA00412.1"  
/db\_xref="GI:216871"  
/translation="PGAQIQHKSIVGVQALPLVMDPATTKAV"  
101..104  
/note="SD sequence for camA"  
115..1383  
/gene="camA"  
115..1383  
/gene="camA"  
/gene="camA"  
/codon\_start=1  
/transl\_table=11  
/product="NADH-putidaredoxin reductase"  
/protein\_id="BAA00413.1"  
/db\_xref="GI:912426"  
/translation="MNAANDNVIVGTGLAGVEVAFGLRAGMGENTRLVGDATVPIPH  
LPLSKAYLAGKATAESLYLRTPDAYAAGNIOLIGTQVATINRDRQVILSDGRALD  
YDRLVLTATGGRPRPLPYASGAVKANNFRILRTLEDEACIRROLIADNRLVITGGYT  
GLEVAATAIKANHATVLLDTAARVLEVTAPVSAFVETLHREGVDIRGTQCGFE  
MSTDQOKVAIVLCEIDGTRLPADLVIAIGILPNCELASAGLQVNDINGVINEHMQTSD  
PLIMAVGDCARFHSQLYDRWRVRIEVPNALQOARKIAAIIICGKVPREDAPMPMSDOY  
EILKMWGLSEGDRITIVRGSIAQDPFSVFLQGDRLAVDTVNRPRPEFQNSQIITD  
RLVEPNILGDESVPLKEIIAAKAKELSSA"  
1427..1430  
/note="SD sequence for camB"  
1439..1762  
/gene="camB"  
1439..1762  
/gene="camB"  
/gene="camB"  
/codon\_start=1  
/transl\_table=11  
/product="putidaredoxin"  
/protein\_id="BAA00414.1"  
/db\_xref="GI:216873"  
/translation="MSKVYVVSHTGTREILDVADGVSLMQAAVNGIYDIYDGGGSA  
SCATCHYVNEAFTDKVPANEREICMLBCTVTAELKPNRLCCQIITPPLDGIYVDY  
PDRQW"  
1442..1759  
/gene="camB"

BASE COUNT /product="putidaredoxin mature peptide"  
ORIGIN 450 a 555 c 575 g 370 t  
Query Match 6.6%; Score 92; DB 1; Length 1950;  
Best Local Similarity 100.0%; Pred. No. 1.6e-09;  
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1311 ccgggtgcccaatcagcaagaagcgcatcgtaagcggtgcaagcactccctctg 1370  
|||||  
Db 1 CCGGGTGCCTCAGATTCAACAGAGCGGCGATCGTCAAGCGCGTCAAGCCTCTCTG 60  
Qy 1371 gctcgatccggcgactaccacaagcgata 1402  
|||||  
Db 61 GTCTGGGATCCGGCGACTACCAAGCGGTATA 92

Search completed: October 4, 2000, 06:41:30  
Job time: 13125 sec

mal\_peptide





GenCore version 4.5  
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2000, 03:16:31 ; Search time 101.52 Seconds  
(without alignments)  
3455.174 Million cell updates/sec

Title: US-09-246-451-1

Perfect score: 1402  
Sequence: 1 ctgcagatcgttaccgct.....ggcactaccagaacggtata 1402

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N.Geneseq-36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1402	100.0	1578	X19926	Pseudomonas putida
2	652.2	46.5	1356	X19916	Bacterial and mamm
3	140.8	10.0	110000	V30458_4	Continuation (5 of
4	140.8	10.0	110000	V30459_4	Continuation (5 of
5	113	8.1	15079	Q91580	S. clavuligerus cl
6	110.6	7.9	5821	T58686	DNA encoding S. fr
7	97.6	7.0	1233	O61452	Cytochrome P450 SC
8	88.2	6.3	1879	Q11126	Sequence encoding
9	83.4	5.9	1194	T58555	Streptomycies prist
10	83.4	5.5	4496	T58553	Streptomycies prist
11	77.8	5.5	53789	V21187	Amycolatopsis medi
12	76.6	5.5	2168	Q73674	Mycinamycin IV hyd
13	74.6	5.3	1998	Q11127	Sequence encoding
14	73.6	5.2	1700	Q45569	Sequence soyC and
15	68.6	4.9	1400	Q14548	Eryf gene. New 6-d
16	58.8	4.2	6085	T70153	S. longisporoflavus
17	44.4	3.2	8169	V26609	Actinomadura hibis
18	40.4	2.9	35089	V27112	Adenovirus 17. Rec
19	39.8	2.8	503	Q14549	Eryf gene recognit
20	39	2.8	1239	N91576	Sequence of clone
21	39	2.8	1239	Q43439	Lol PI cDNA clone
22	38	2.7	1929	V57472	Sorghum bicolor (L
23	37.2	2.7	1000	V63929	Mycobacterium tube
24	37	2.6	13987	T80415	Orotidine-5'-monop
25	36.8	2.6	13987	T80415	Hybrid srmg/tylg O
26	36.8	2.6	43280	T80413	Tylosine synthase
27	36.6	2.6	1747	T12248	Cytochrome P450 is
28	36.4	2.6	8478	Q81792	B. subtilis biotin
29	36.4	2.6	29879	Q46806	Erya region of S.
30	35.8	2.6	1164	Q72603	Plant insecticidal
31	35.8	2.6	1286	V74187	Barley pathogen in
32	35.8	2.6	1286	X23274	H. vulgare cystein
33	35.8	2.6	3076	X23275	H. vulgare cystein

C	34	35.2	2.5	750	1	X32469	P. fluorescens mal
	35	35	2.5	1485	1	Q87721	Human auxiliary cy
	36	35	2.5	1485	1	Q87722	Human auxiliary cy
	37	35	2.5	1485	1	T17409	Human derived cyto
	38	35	2.5	1485	1	T28387	Human cytochrome P
	39	35	2.5	1485	1	T28388	Human cytochrome P
	40	34.6	2.5	9960	1	V58939	Mycobacterium smeg
	41	34.2	2.4	930	1	X25146	Wheat Type I gluta
	42	34	2.4	716	1	V49535	harkiri cDNA. New
	43	34	2.4	954	1	Q66592	GST-II-27 gene. PI
	44	34	2.4	4682	1	Q34557	GST-II-27 cDNA. 27
	45	33.8	2.4	1564	1	Q38955	Lipase gene #1. DN

ALIGNMENTS

RESULT	1
ID	X19926
AC	X19926; standard; DNA; 1578 BP.
DF	11-JUN-1999 (first entry)
DE	Pseudomonas putida cytochrome P450 protein P450-cam encoding DNA.
KW	Bacterial; mammalian; cytochrome P450; chimeric; fusion protein;
KW	oxidise; hydrocarbon; carbon-hydrogen bond; hydroxylating;
KW	bioremediation; environmental pollutant; ss.
OS	Pseudomonas putida.
PN	W09908812-A1.
PD	25-FEB-1999.
PF	17-AUG-1998; U16979.
PR	20-AUG-1997; US-056754.
PA	(UYRP ) UNIV ROCHESTER.
PI	Jones JP, Shimoji M;
DR	WPI: 99-190131/16.
DR	P-PSDB: Y04128.
PT	New P450 fusion proteins - comprising a portion of a bacterial
PT	cytochrome P450 protein and a portion of a mammalian cytochrome P450
PT	protein
PS	disclosure; Page 13-14; 51pp; English.
CC	The present invention describes a fusion proteins comprising a portion
CC	of a bacterial cytochrome P450 protein and also a portion of a mammalian
CC	cytochrome P450 protein. The fusion protein can oxidise hydrocarbons or
CC	any compound having a carbon-hydrogen bond. The fusion protein can be
CC	used for hydroxylating a compound to be oxidised. It can also be used in
CC	the bioremediation of an environmental pollutant. Since the fusion
CC	protein is soluble, it can be subject to structural elucidation by X-ray
CC	crystallography for designing functional proteins. It can be readily
CC	expressed in soil bacteria to facilitate bioremediation. The present
CC	sequence encodes Pseudomonas putida cytochrome P450 protein P450-cam
CC	from the present invention.
SQ	Sequence 1578 BP; 350 A; 490 C; 436 G; 302 T;

Query Match 100.0%; Score 1402; DB 1; Length 1578;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1402; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ctgcagatcgttaccgctgagcagatcgtatcacccagcgttttccatcagcagagcc	60
DB	1	CTGCAGATCGTTATCCGCTGCGATCTGATCACCCGCTTTTCATCGACGAGGCC	60
QY	61	agcaagacactgaactggtcaagcagagcagcactgataaacccgtagatgactcact	120
DB	61	AGCAAGGACCTTGAACCTGTGAAGGACGACCTGATCAACCGTAGATCGACTCCACT	120
QY	121	ctttagcaaacccggttccagagagaacaacaacatgacactgataaaccaataaagca	180
DB	121	CTTTAGCAACCCGCGTTCCAGGAGAACACAAATGACATCGAAGCAACTACAAAGCA	180
QY	181	agcccaatctgcccctctgcaccccaatgtgcagagcactggtatctgcagatcgaca	240
DB	181	AGCCCAATCTTGCCCTCTGCGACCCCATGTGCCAGAGACACTGTGATTCGATTCGACA	240

```
QY 241 tglacaatccgctgaatctgtctgcgagctgcaggagccctggagcaatctgcagaat 300
    |||
Db 241 TGTACAAATCCGTCGAATCTGTCCGGCTGCAGAGCCCTGGCGAGTTCTGCAGAAAT 300
QY 301 caaagctaccgagctctgtgtgtgacctgcctgcaacgagcgagcaactgtatgcacatccgcg 360
    |||
Db 301 CAAAGCTACCGGATCTGTGTGTGACTGTGCAACGGCGGACACTGTGATGCCACTCGCG 360
QY 361 gccaactgtatccgtgagccttaagaattaccgcaactttccagcgagtgccggtta 420
    |||
Db 361 GCCAACTGATCCCTGAGGCGCTATGAAGATTACGCCCACTTTTCAGAGATGCCCCGTTCA 420
QY 421 tccctcgtgaagcgcgagccttaagactlccacccctgatgagatgcgcccagc 480
    |||
Db 421 TCCCTCGTGAAGCGCGGAGGAGCCTACGACTTTCACCTCCAGTGCATGCGCCGAGC 480
QY 481 aggcgcagttctgtgcctgcgcgaacaaagtgtgtgcatgcccgtgtgtgataagtctg 540
    |||
Db 481 AGGCGCAGTTCTGTGGCTGGCCACCAAGTGTGTGCAAGCGGTGTGATTAAGCTGG 540
QY 541 agaacgcgataccagagcgtgcctgtcgtctgaatcgaagacctgcgcgcgcaagagacgt 600
    |||
Db 541 AGAACCGGATCCAGAGACTGTGCTGCTGCTGATCCAGAGCCTGCCCCCAAGAGACGT 600
QY 601 gcaacttcaacgagactacgacccgaacccctlcccgatagcagatcttcatgtctgcag 660
    |||
Db 601 GCAACTTTCACCGAGGACTACGCCGAACCCCTCCGATACCGATCTTCATGCTGCTGCAG 660
QY 661 gtctaccggaagaagatatccgcgaattgaataactaaagatcagatgacccgctcgg 720
    |||
Db 661 GTCTACCGGAAGAATATCCGCCACTTGAATACCTTAACGGATGAGATACCCGCTCCG 720
QY 721 atgagcaatgactctgcagagagcgagagagcgctcgaactctctgataccgata 780
    |||
Db 721 ATGAGCAATGACTCTGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 780
QY 781 tcgagcaacgcaagcagaagcgcggaacgcgacgtatcagacatcgtgtgcgaacgagcag 840
    |||
Db 781 TCGAGCAACGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 840
QY 841 tcaatggggcgacccgataccagatggaagcaagagagatgtgtgaccttactgtcg 900
    |||
Db 841 TCAATGGGGCGACCGATCACCAAGTACGAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 900
QY 901 ggcgcctgtgatacgtgtgtaaatcttccatgactcagatcgagaggttccctgcgcaaaagc 960
    |||
Db 901 GCGCGCTGATACGCTGTGTAATTTCTCTCAAGCTTCAAGATGGAGTTCTTCCGCAAAAAGCC 960
QY 961 cggagcagtcgcaagagctgtatcgaagcgtlcccgagcgatltccagccgcttgcgagaaac 1020
    |||
Db 961 CGGAGCAGTCGCCAGAGCTGATCGAGCGTCCGAGCGTATTTCCAGCGCTTGGGAGGAGAAC 1020
QY 1021 tactcggcgctctcgtcgtgtgtgcgagatggcgcatctcgaactcgaatgaagtctc 1080
    |||
Db 1021 TACTCGGCGCTCTCTGCTGTGTCGATGGCGATGCCGATCTCCAGCTCCGATTAAGAGTTTC 1080
QY 1081 atgagcgtgcaactggaagaagagtgacacaga tctctgtacaccgagatgagctgtcggagcttg 1140
    |||
Db 1081 ATGAGCGTGCACACTGGAAGAAAGGTGACCAAGATCTCTGCTACCGCGAGATGCTGTGCGCTGG 1140
QY 1141 atgagcgcaaaaaagcgtcccgatgcaagctgcaactcagtcgcaaaaagtttccacaca 1200
    |||
Db 1141 ATGAGCGCAAAAAAGCGCTGCGATCGATCGATCACTTCCGCAAAAAGTTTCAACACA 1200
QY 1201 ccaactttggcagcagcagcagcatctgtgcttggcagacacctggccgcgaggaataca 1260
    |||
Db 1201 CCACCTTTGGCCACGGCAGCAGCATCTGTGCTTGGCCAGCACTGTGCGCGCGGGAATCA 1260
QY 1261 tcgtacccctcaagagatgagctgacagagatctcgtactctcattgcccgggtgagccc 1320
    |||
Db 1261 TCGTCAACCTCAAGAGATGCTGACCAAGATTTCTTCACTTTCTCATTTGCCCCGGGTGCC 1320
QY 1321 agattcagacaagagcgcgcatcgtcagcgctgcagcgactccclctgtctggatc 1380
```

```
Db 1321 AGATTGAGCACAAGACGGGATGTCAGCGGCTGAGGACGATCCCTGTGTGGATC 1380
QY 1381 cggcgactaccaaagcgata 1402
    |||
Db 1381 CGCGACTACCAAGCGGATTA 1402

RESULT 2
X19916
ID X19916 standard; DNA; 1356 BP.
AC X19916;
DT 11-JUN-1999 (first entry)
DE Bacterial and mammalian chimeric cytochrome P450 protein encoding DNA.
KW Bacterial; mammalian; cytochrome P450; chimeric; fusion protein;
   oxidase; hydrocarbon; carbon-hydrogen bond; hydroxylating;
   bioremediation; environmental pollutant; ss.
OS Synthetic.
PN MO9908812-A1.
PD 25-FEB-1999.
PE 17-AUG-1998; U16979.
PF 20-AUG-1997; US-056754.
PA (UKRP ) UNIV ROCHESTER.
PI Jones JP, Shimoi M;
DR MPI; 99-190131/16.
P-PSDB: Y04126.
PT New P450 fusion proteins - comprising a portion of a bacterial
   cytochrome P450 protein and a portion of a mammalian cytochrome P450
   protein
PS Clam 10; Page 5-6; 51pp; English.
CC The present sequence encodes a fusion proteins comprising a portion of a
   bacterial cytochrome P450 protein and also a portion of a mammalian
   cytochrome P450 protein. The fusion protein can oxidise hydrocarbons or
   any compound having a carbon-hydrogen bond. The fusion protein can be
   used for hydroxylating a compound to be oxidised. It can also be used in
   the bioremediation of an environmental pollutant. Since the fusion
   protein is soluble, it can be subject to structural elucidation by X-ray
   crystallography for designing functional proteins. It can be readily
   expressed in soil bacteria to facilitate bioremediation.
SQ Sequence 1356 BP; 349 A; 386 C; 323 G; 298 T;

Query Match 46.5%; Score 652.2; DB 1; Length 1356;
Best Local Similarity 99.5%; Pred. No. 2e-163;
Matches 654; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 156 atgagcactgaaacacatacaagaacgcgaactctgccccttgcacccaatgtgca 215
    |||
Db 1 ATGAGCACTGAAACCATACAAACGCAACGCCAATCTTGCCTTGCCACCCCATGTGCCA 60
QY 216 gagcactggtatgactcgacatcgatacaatcgtgaaatctgtcggcggtgcag 275
    |||
Db 61 GAGCACCCTGGTATTTCACATTTCGACATGTACAAATCCGTGCAATTTGTTGCGGCGTGCAG 120
QY 276 gaagccttggcagttctgcaagaatacaaacgtlacccgagatctgtgtgagactcgtgcaac 335
    |||
Db 121 GAGGCGTGGGCACTTGTGCAAGAATCAAAACGTAACCGGATCGTGTGAGACTCGCTGCAAC 180
QY 336 ggcgcgaactgtatgcgcactcgcgcgcgaactgatactcgtgagagctatgaagattaccgc 395
    |||
Db 181 GGGGGAACACTGAGTCCCACTGCGGCGCAACTGATCCGTGAGGCGCATGAAAGTTTACCGG 240
QY 396 caatttccagagagtcgcccgttcatccctcgtggaagcgcggaagcctacgactcatt 455
    |||
Db 241 CACTTTTCCAGCAGTGTGCCGTTTCAATCCCTCGTGAAGCCGCGCAACGCTTACATTT 300
QY 456 cccaactctgatatcgcgccgagcagcgccagtttctgtgcgttggcacaacaaagtgtt 515
    |||
Db 301 CCACACTCGATGATGATCGGCCGAGGAGGCGCACTTTCGTGCGTGGCCAAACGAAGGT 360
QY 516 ggcatacgcggtgtgtgataaagtgtgagaacgcgataccagagagctgagctgcgtgatac 575
    |||
Db 361 GGATGCGCGGTGTGATTAAGCTGGAAGAACCGGATCCAGAGAGCTGCGCTGCTGATATC 420
```

Qy	576	gagagcctcgcgcccgcgaagaacagtgcacacttcaccgagagactaagccgaaccccttcgcg	635
Db	421	GAGAGCCTTCGCCGCCGAAGACAGTGCACCTTCACCGAGAGACTACCGCAACCCCTTCCCG	480
Qy	636	atagcgcattctatgtctgtctgcaggtctaccgcggaagaatatcccgcaacttgaataac	695
Db	481	ATAGCGATCTTCATGCTCTGCTCGCAGGTCACCGGAAGAAGATATCCGGCACTGGAATAC	540
Qy	696	ctaaagagatcagaatgaccgcctcgatgacagatgacagacttcgcagagccgaagagggcg	755
Db	541	CTAAAGGATCAATATACCCGTCGCGGATGGACACATATACCTTCGCAAGAGGCCAAAGGAGGCG	600
Qy	756	ctctcgcacatctatataccgatactcatctcgacgaacgcaggcagaagcccggaaccgac	812
Db	601	CTCTACGACTATCTCATATACCGATCATCTCGAGCAACGACGACAGACCCGGGAATGAAC	657

RESULT 3  
 V30458\_4  
 Continuation (5 of 6) of V30458 from base 400001 (Rhizobium species plasmid pNGR234a. )  
 WP Sequence Split into 6 fragments LOCUS V30458 Accession V30458  

WP	Fragment Name	Begin	End
WP	V30458_0	1	110000
WP	V30458_1	100001	210000
WP	V30458_2	200001	310000
WP	V30458_3	300001	410000
WP	V30458_4	400001	510000
WP	V30458_5	500001	534720

Query Match	10.08;	Score 140.8;	DB 1;	Length 110000;
Best Local Similarity	48.58;	Pred. No. 1.6e-27;		
Matches 478; Conservative	0;	Mismatches 502;	Indels 6;	Gaps 3;

[illegible]

Dd	47322	GAGAGGAATGTCGGCGGCATCGGTCTTTTCTTGTCGGGGGCTCCACACGATTGCC	47381
Oy	921	aatttcctcagcttcacgatgagttcccttgcccgaaggccggagcatcgccagaactcy	980
Dd	47382	GCCCCCATTTGGCTTTGACATB6CCTAATTCTTGACACGCAATCCAAGCACCGAGTTGTTG	47441
Oy	981	atcgagcgltccgcgagcglatlccaccgccgtctlgcgaggaactacctcgcgacttccgtcy	1040
Dd	47442	CGAACGAAACCGGCTCGGGCTCGGGCTCGCCGCTGAGSAACTGCTGGCGGCTTAATTAACC	47501
Oy	1041	gttgcgcagtgcgcgcatacctcaacctccgatlacagagittcatgtgcbgtgcactgaaana	1100
Dd	47502	GTTCCGATTAATCCCGCGTGGCCACGAAGAACAATCGAGTTCCNAAGCGTGCCTATCCGTAG	47561
Oy	1101	ggtgcacgagalcctctctaaccgcgagatctgtcttgcccttgtatgagcgcgaanaacgcctcy	1160
Dd	47562	GGTGATTTATGTTTTCTGTCGCCGCGCATGTTGCTATATGAGAAATCCGTCGATAATTTAAGTGC	47621
Oy	1161	cgcagtcgcagtcgaacttcagtcgcgaagaagtlclcacacaaccaacctlltbgccaacycagc	1220
Dd	47622	CCCAACACTATTCGATTCGTGGCAGCACAGGATAAACGACGACACGCCCTTTGGCTATGGACCA	47681
Oy	1221	catctgtgcctcttgcgcagcaccttgyccgcgcgggaatacatgtlcaacctcaaagaaatgy	1280
Dd	47682	CACCTTTGCCACGGAGGCGCATCTGCAACGGCGAGAAATCGCTATTTGGCTACTGTAATGG	47741
Oy	1281	ctgaccagagatlctctgaacttccat	1306
Dd	47742	TTAGCGCGCATCCACAGCTTTCGGCAT	47767

## RESULT 4

Continuation (5 of 6) of V30459 from base 400001 (Rhizobium species symbiotic plasmid						
WP	Sequence split into 6 fragments	LOCUS V30459	Accession	V30459		
WP	Fragment Name	Begin	End			
WP	V30459_0	1	110000			
WP	V30459_1	100001	210000			
WP	V30459_2	200001	310000			
WP	V30459_3	300001	410000			
WP	V30459_4	400001	510000			
WP	V30459_5	500001	531655			
<hr/>						
Query Match	10.0%	Score 140.8;	DB 1;	Length 110000;		
Best Local Similarity	48.5%;	Pred. No. 1.6e-27;				
Matches 478; Conservative	0;	Mismatches 502;	Indels 6;	Gaps 3;		
OY	324	actgcgtcgaacgaggagacaatgatgcacactcgccggccaactgatactcgtagacctat	383			
Dd	46785	ACGCCCGATGAGCAGGGTACTGGGTGATTAAACGGGGGAGAGACAAGCCCGGTCTTT	46844			
OY	384	gaagaattaccgcaaatcttcacgagaygccccyttaacccctgtgaagccggcgaaagcc	443			
Dd	46845	GAGGATTACAGAGACCTTTTCACGCCATGCAGACATCTTGCTTGCGCACTTGGGAACAC	46904			
OY	444	tacgacttattcccaccactcgaltgatccgcgcgagcacgagccagtttgtgcgtgcc	503			
Dd	46905	TGCCCCGTCATCCGCCCTTGAGATTGGACCACC GCCGCCCATGTGAGTGTTCGTGACACTCTA	46964			
OY	504	aaccaagtgtgtgacacgacgcggtgtgtgataagctbgagaaacggatccaggaagctggcc	563			
Dd	46965	AATCTCTGTTCTCATCCAGGCGG6TGTGGCGCTGGAGCCGCACTATCCATGCCAGACA	47024			
OY	564	tgtctgcgtcatccgagagct---gcgccgcaagagacafgcaactlcaaccgaggaactac	620			
Dd	47025	G6CGCGTCACTGACTGCTGATTGCGAAAGAGAACAACACTGTGACGTGATGAAGATTTC	47084			
OY	621	gccgaacccctcccgatccatctcatalgtgtgtgtcgcgaggtctaacgcggaagaatatcc	680			
Dd	47085	GCCTTGGCCCTTTACCTTAGCGTTTTCTT-CAGCTTTCTGGGACTTTCCAGAGCGCATC	47143			
OY	681	cgcacatcgaaataacctaacgatacgaatgaacccgtccgagtaggcagatgaacctcgca	740			

Db	47144	CGAAGTACTTGTGGGCTGGGTAAAGCGATTTCGTCACCGCAACCGCAAGAAACGAAAG--	47201
Qy	741	gaagccaaaggagcgctctaagactatctgataccgaatcatcgacaacgcaaggagaag	800
Db	47202	GCAGAGGCCCGCGTCGGTGGCGCTTTATATGACGAAATGGACGCATCGCGCGCAAGTCG	47261
Qy	801	ccggagacccgaagcgtatcaagactgtttcccaacggccaggtcaatgaggcaagctaac	860
Db	47262	CCAGCTGTGCGATTTCATGACCTTCGTCTGTCACGGCGAAATGAGAGGCCGCTCTTAACA	47321
Qy	861	agtgacgaagccaaaggatgctgtgacctgtactgttcgagcgcgcttgatacgttgctc	920
Db	47322	GAAAGAGGAAGTCCGGCGGATCGGTGTGCTTTCTTGTCGCGGGGCTCGACACGGTGGC	47381
Qy	921	aattccctcagcttaagatgtagtctctctgtgcccacaagccggagacatcgcgaagagctg	980
Db	47382	GCCGCCATTGGCTTTGACATGGCCATCTCTTGACGCCAATCCAAAGCACAGAGTTGTTG	47441
Qy	981	atcgagcgctcccgagcgatctatcccaagcgcttgcgaggaactactccggcgcttcgcctg	1040
Db	47442	CGAAGCGAAGCCGGCTCGGCTCGGCGCTCGCGTGGAGAACTCTCTCGCGCCATTCAAC	47501
Qy	1041	gttcgcgaatgycgcgaatcctcaactccgaattacaagagtttcaatgycgtgaactagaaga	1100
Db	47502	GTTTCGATAATATCCGCGTGCGCCGAGAGACATCGAGTTTCGAAGCGGTCCATCCGTGAG	47561
Qy	1101	gctgacgaagatcctgtactaccgagaatgctgtctgacctgagatgagcggaaaaacgctgc	1160
Db	47562	GGTGATTATGTTTCTCTGTCCCGCGATGTGCTTAATCGAGATCCGTCTGAATTTAAGTGC	47621
Qy	1161	ccgagtcacgtccgaacttaagtcgccaagaagtttacaacaccactcttggcacaagcgagc	1220
Db	47622	CCCAACACTATGATCTGTGGACGACAGATACACGACACACGCGCTTGGGTATGAGACA	47681
Qy	1221	catctgtgcttggcgaagactgtgcccggggagaatcatcgttcaacctcaaggaatg	1280
Db	47682	CACCTTTCCACGAGGCGCATCTCCACGGCGCAAAATCGTATTGGCTTACGTGAATGG	47741
Qy	1281	ctgaccagaatctcgaacttccat	1306
Db	47742	TTAGCGCGCATCCACGCTTCCGCAAT	47767
RESULT	5		
ID	Q91580	Q91580 standard; DNA; 15079 BP.	
AC	Q91580;		
DT	13-NOV-1995	(first entry)	
DE	S. clavuligerus cla gene region.		
KW	clavulinic acid; clavulinate; antibiotic; beta-lactamase-inhibitor;		
OS	Streptomyces clavuligerus.		
CH	key	location/Qualifiers	
FT	cds	complement (49..1745)	
FT	FT	/*tag= a	
FT	FT	/label= ORF1	
FT	FT	2216..3937	
FT	FT	/*tag= b	
FT	FT	/label= ORF2	
FT	FT	3940..5481	
FT	FT	cds	
FT	FT	/*tag= c	
FT	FT	/label= ORF3	
FT	FT	5634..6595	
FT	FT	cds	
FT	FT	/*tag= d	
FT	FT	/label= ORF4	
FT	FT	/note= "cla gene"	
FT	FT	6611..7588	
FT	FT	cds	
FT	FT	/*tag= e	
FT	FT	/label= ORF5	
FT	FT	7895..9076	
FT	FT	cds	
FT	FT	/*tag= f	
FT	FT	/label= ORF6	

FT	cds	9241..10998	
FT		/*tag=	g
FT		/label=	ORF7
FT	cds	complement (10998..12296)	
FT		/*tag=	h
FT		/label=	ORF8
FT	cds	complement (12662..13365)	
FT		/*tag=	i
FT		/label=	ORF9
FT		13769..14995	
FT		/*tag=	j
FT		/label=	ORF10
PN	CA2108113-A.		
PD	09-APR-1995.		
PE	08-OCT-1993;	108113	
PR	08-OCT-1993;	CA-108113.	
PA	(UTAL-) UNIV ALBERTA.		
PI	Aldoo KA, Jensen SE,	Paradkar AS;	
DR	WPI; 95-207301/28.		
DR	P-DSDB: R77858;	R77859;	
DR	R77865;	R77866;	
PT	Clavulanic acid biosynthesis enzymes and corresp. DNA - useful for	R77860;	R77861;
PT	biosynthesis of the antibiotic in Streptomyces hosts which do not	R77862;	R77863;
PT	naturally produce clavulanate	R77864;	
PS	disclosure; Fig.2; 41pp; English.		
CC	A library of S. clavuligenus NRRL 5741 genomic DNA fragments (15-22		
CC	kb) constructed in cosmid pLAFR3 was screened for the cla gene using		
CC	a probe based on a partial N-terminal sequence from the CLA enzyme.		
CC	Isolated clone 66L1 included a 15kb fragment having the sequence		
CC	given in Q91580 that included the cla gene (ORF4).		
SO	Sequence 15079 BP;	2097 A;	5488 C;
			5324 G;
			2170 T;

Query Match	8.1k:	Score 113:	DB 1:	Length 15079:
Best Local Similarity	51.3k:	Pred. No. 1.9k-20:		
Matches 291:	Conservative 0:	Mismatches 270:	Indels	Gaps 1:
OY 736	tcgcgaagacgaagaagcgctctacagatctctgataccgatacgaagaagcgc	795		
Db 14340	TCGCCAAGGCCCCGGAGCAGACTTCGACGECTTCTCCGGGACTGGTTCAGGAGCGGATCG	14399		
OY 796	agaagcgggaagacgcagcgtatcagcatcgtctgcgaagcgcaggtcatctgycgaccca	855		
Db 14400	AGAACCCGGGACCGACGACTGTATACGCCCGGTCGTCATGAGCAGAGTGCGGGCGGCATC	14459		
OY 856	tcaccagtgaacgaagccaagaagatgtgtgtcctgttaactgtgtcggcgctgtatacgg	915		
Db 14460	TGCGGGTGAAGAGATGCTCCCGATGTCGCCGCTGCTCTGTGGCCGCTACCGCGACCA	14519		
OY 916	tggtcaattcttcagcttcagcatcagatgagttcttcggccaagaagcccgagatcgcgaag	975		
Db 14520	CCACCAAGCAGGCGAGCTTGAGCCTGCTCAGCCTCTACTACCGACCCGGAGCTGGCCGGCC	14579		
OY 976	agctgatcgaagcgttcccgagcgtatctccagccgctcttgaggaactaactccgycgttct	1035		
Db 14580	GCTTCACCGAGACCCCGCCCTGCTGCCCCAAGCGCGTGAAGAGAGCTGCTGCGTTCCACT	14639		
OY 1036	cgctgtgtgcgcatgagccgcacatctcctaactccgattacg-----agttcaatgagcgtgc	1089		
Db 14640	CCATCGTTCAGAACGGGGCTGGCCCGTGCCGGGTGAAGAGCTCAGTCCACGATGTGC	14699		
OY 1090	aactgaagaagaagtgcaccagatcctctgtcagcagagatgctgtctcgcgtctgatatgagcg	1149		
Db 14700	TCAATCCGGGCGGGGAGGGCTGGTGGCTGTGCTGCGCGGSCAACCGGGACAGACGAGG	14759		
OY 1150	aaaagcgtccgcgcatgcaagctgcactcagtcgcgaagaagytllcacaaccaacttgg	1209		
Db 14760	TCTTCCCGGACCCGACCGGGTGGACGTGAGACCCGAGACCCGCCGCCCATCTCCGCTTGC	14819		
OY 1210	gcacagcagcatalcgtccttgcgcagcacaactcgtccgcgcggaatatcgtctacc	1269		
Db 14820	GCCACGGCATCCACAGGCTGCGGCGAGTGGCTGGCCCGGGGTGAAGTGAAGAGATTC	14879		

Oy	1270	tcaagaa	tgtcgt	accaggat	ctctg	1296
Db	14880	TCGCCG	CGGTCT	GCCTG	GATGCCC	14906

RESULT	6
ID	T58686/c
AC	T58686 standard; DNA; 5821 BP.
DT	17-FEB-1998 (first entry)
DE	DNA encoding S. fradiae tylosin biosynthesis gene products.
KM	Streptomyces fradiae tylosin biosynthetic gene product; production; antibiotic; ds.
OS	Streptomyces fradiae.
FH	Key
FT	Location/Qualifiers
CDS	complement(73..1260)
FT	/tag= a
FT	/product= tyle
FT	complement(1277..1957)
CDS	/tag= b
FT	/product= tylD
FT	complement(2286..4049)
FT	/tag= c
FT	/product= tylH
FT	4094..4927
FT	/tag= d
FT	/product= tylF
CDS	complement(5046..5660)
FT	/tag= e
FT	/product= tylJ
PN	US5672497-A.
PD	30-SEP-1997-A.
PE	21-DEC-1995; 575843.
PR	12-MAY-1989; US-351350.
PR	21-MAR-1986; US-842330.
PR	25-JUL-1986; US-890670.
PR	24-FEB-1987; US-018237.
PR	06-AUG-1991; US-742222.
PR	28-JUL-1993; US-107232.
PR	17-FEB-1994; US-198672.
PR	21-DEC-1995; US-575843.
PA	(ELIL ) LILLY & CO ELI.
PI	Cox KL, Fishman SE, Hersberger CL, Seno ET;
DR	WPI: 97-488860/45.
DR	P-PDOB: W33272, W33273, W33274, W33275, W33276.
PT	DNA encoding Streptomyces fradiae tylosin biosynthesis gene products
CC	- for increasing tylosin production in Streptomyces spp.
CS	claim 1; Columns 21-38; 38pp; English.
CC	The present sequence encodes Streptomyces fradiae tylosin
CC	biosynthetic gene products, useful to increase the production of
CC	the antibiotic tylosin in Streptomyces spp.
CO	Sequence 5821 BP; 799 A; 2066 C; 2087 G; 863 T;

Query Match	7.9%	Score 110.6	DB 1	Length 5821
Best Local Similarity	47.4%	Pred. No. 6.1e-20		
Matches 397	Conservative	0	Mismatches 434	Indels 6
			Gaps	2
QY 466	tgaatccgcgccgagcagcgccagttctgtgctctgccaacccaagtgglttgcatgcccgg	525		
Db 3493	TGGAACCCGCCGACACGAGGCGCTCCGGCCACTTCATCCCGAGTTCGGCGTGGCGC	3434		
QY 526	tgtgtgataaagcttgagaagaacggatccagaggagctgtgctgtctgatacgagaagcttgc	585		
Db 3433	GGGTGCGGGAGACTCGGCCCTTCGTCGAACAGATGTGTCACCGGCTGTGTGATGACCTCA	3374		
QY 586	gcccgcaagg--acagttgcaacttccacggagctacgcgcgaaccccttccgatacgca	642		
Db 3373	CCGCCCGCGGAGACAGGCGGAGTGTGCTGTGGCGAATTGGCGCTGCCCATGGCGACCCAGG	3314		
QY 643	tcttcatagtctctcgcaaggtctaccggaagaagatatcccgcaacttgaataactcaagc	702		
Db 3313	TGATATGCCCGCTGTCCACATTCCTCTACAGAGACGGGAGTACTTCCAGAGGCCACCG	3254		

OY	703	atcagatgacccgtctcggatgtagcaatgaacttcgcaagagccaaagagcgctctacg	762
Db	3253	AACAGGCCACCGCCCGCGCGCGCGCGAGAGCGCTGGAGGCGCTGCGAGAGCTCGCGG	3194
OY	763	actatctgatatccgatcatctcgaagaaacgcgagaaagccggaaaccgagacgctatcaaga	822
Db	3193	ACTAATCTCGAACCGGCTATATCAGCGGCGAAGACCGCGCGGAATCCGGCGACGGGATGCTGG	3133
OY	823	tcgtgtccaaacgagcaggtlcaatlggagcagatcaaccaagltgcgaagagcgaagaaagatgt	882
Db	3133	GCAGAGATGATGGCGCCAGGCGCCGGGCTGGCTGTGACGCGCCGACGCTGCGAGCAACG	3074
OY	883	gtgagcctgttaactgctcgagcgccctgtgatacgtgtgtlcaatttcctcagatctcaagatg	942
Db	3073	CGGAGCTCTGCTCGCGCGCGGCGACAGACCGCGCGCAATGAGTCAAGATGAGCGTGC	3014
OY	943	agttcttcggccaaagacccggagagatctgcagagaagctlgtacgaagcgctcccgagcgatlc	1002
Db	3013	TGCTTTCTGCTACACACCACCCACGGCGCTCGCGCGCACTACCGTGAATCCCGGCTGTGTGC	2955
OY	1003	cagcgccgttcgagaaactaactccgagcgctcttcgtgtgctgcagatg--gcgcatacc	1055
Db	2953	CGGGCGGGGTGAGACGACTGTGGCTGATGTGTGATGCGCGACGGGCTGCGCGCTGCG	2894
OY	1060	tcacactccgattacgagttcatatggtcgctgaacttgaaagaaagttgacagatctcgtctac	1119
Db	2893	CCACCGCGGACATGAGATGATGACGGGCGCACACATCCGCGCGCGCGCGACTGAGTGTCTCC	2834
OY	1120	cgaagatgctgtctgagccttgatgtagcggaaagccgtccgcgaatagacagctcagacttca	1179
Db	2833	TGCTTGCCGCGCCACACCGGAGACAGGCGCGCTTCTTCTCCGACGGAGGCGCTTTCGATATCC	2774
OY	1180	gtcgccaaagaglttccacacacacacacttlttgccagcgacacatctgtgtcctltggcagc	1238
Db	2773	ACCGGATCGCGAGGCGCGCATGTGCGCTTGCGGTACGAGACCCACGATGCTGCGGAGAA	2714
OY	1240	acctggtccgcgggaatacatcgtcaacctcaagaatagtgtcgtacagagatctctg	1296
Db	2713	ACCTGCGCGGATGGAATGGAATGTGGCCCTTCGCGCGCGTGTGAGAGAGGCTGCGCG	2657

```

RESULT      7
NC_061452   061452 standard; DNA; 1233 BP.
CC          061452;
CT          24-NOV-1994 (first entry)
C            Cytochrome P450 SCA-2 gene.
C            Cytochrome P450 SCA-2 gene.
C            Cytochrome P450 SCA-2; Streptomyces carbophilus;
C            treatment of hyperlipidaemia; drug preparation; ss.
C            Streptomyces carbophilus.
CS          key location/Qualifiers
cds         1..1233
             /*tag= a
             /product= cytochrome P-450 containing activity

j06070780-A.
15-MAR-1994.
28-AUG-1992: 229969.
28-AUG-1992: JP-229969.
(SANY ) SANKYO CO LTD.
WPI: 94-128679/16.
P-PADB; R51368.
Cytochrome P450 SCA-2 gene - from Streptomyces carbophilus,
useful in treatment of hyperlipidaemia
Claim 3; Page 12-14; 18pp; Japanese.
SCA-2 encodes a protein having cytochrome P-450 activity.
P-450 can be prepared commercially for use in the preparation of a
drug for the treatment of hyperlipidaemia.
Sequence    1233 BP;       207 A;       472 C;       372 G;       182 T;

Query Match Best Local Similarity 46.3%; Score 97.6; DB 1; Length 1233;
Matches 399; Conservative 0; Mismatches 454; Indels 9; Gaps 2;

```

OY	458	cacctcgatgatgacgtccgcaggaagcggccaattctgttgccctggc---	caaccagaatggt	514
Dd	297	CATGGGAGTGGATGCCCCCGACAGCACGGGACGCCTGCCCGCATATCATTAGGAGACTTCAC		356
OY	515	tggcatgcgcgggttggtagtaaaccttgagaaacccgataccaaggagacttgcctgcctgat		574
Dd	357	GGTGAAGGCGATCAAAAGGGCATTCGGCCCCGGAGACGTCAAGCATGCTGCACGGCTTCATGA		416
OY	575	cgaagagcctgcgccgcgaagaacaagtcaactcaaccaaggagactacgcgcgaacctccc		634
Dd	417	CGAATGCTCCCGCCGGGACCACACCGCCGATCTGATCGACGATTTGCCCTCCCGCTCC		476
OY	635	gatacgaacttcacttgtctgtctgcaggtctcacgcggaagaagatalatcccgacttgaata		694
Dd	477	GTCATGCTGATCTTGCCATATGCTCGGCGCTCCCTTAGCCGACACAGATTTCTTCACGA		536
OY	695	cctaagagatcagaatgaacccgtcccgatgtagtcaagcatlgaacttcgcagaagccaagagc		754
Dd	537	CGCGAGCAAGCGCCTGGTGTGACAGGGGGTGGAGGCGCGAAGATGGCCGTGCGCCCGGGACGA		596
OY	755	gctctaagactaatcttgatlaacgatatcagaacaaacgaagcagaagccgcggaaacgcgaacg		814
Dd	597	CTTGAGAGCGCTAACCTGGACAGGGCTGATACACAAGCTGAGTCCGAACCCGGGACGGGCT		656
OY	815	tatacgaactgtgtgccaacgcgaagttcaatlvgtagcgaaccatlatcaagatgaagaagccaa		874
Dd	657	CTCTGGGCAAACTGGTCTACCAACCAAGCTGGCGGACGGCGAGATGACCCGGCGAGCTGAT		716
OY	875	gaagaaatgtgtgacctgttacttgatcggcgcgtgataagctagaatgttactcaactt		934
Dd	717	CTCCACAGCGCCTGCTGCTGCTGCTGCGCGGCTGATWAGACACAGGCGCTCGATGACCTCGCT		776
OY	935	cagaatgagatctccttgccaagaacccgcggaagcatlgccaagagctgataagcgtccoga		994
Dd	777	CAGGCTCATACCCCTGCTCGAACAACCCGACACAGACGCCGCCCTTCGCGCCGACCCGTC		836
OY	995	gcgatlatacagccgccttgccgaagaactaac-----tcggcgcccttcgctgtgttgcga		1048
Dd	837	CTCTGGTCCCGGACCGGTGCGAGAACTGTCGGCGTCTGAGGCATTCGCGGACATCGCGGG		896
OY	1049	tggccgcaatcctcaactccgcgattacgaatlltaatlgcgttgcaacttgaagaagtigacca		1108
Dd	897	CGGCGGATTCGCCACCCCGCACATCGAGATCGAGGACAGACTCATCCGGCGGTGAAG		956
OY	1109	gactcgtctaacgcgaagatgtgtgtcgtgcctgagtagcgcgaaaacgcctgcgcgatga		1168
Dd	957	AGTAATGCTACCAACTTCATTCGCGCAACCGCAGACTTCGGTGTGAGAACCCGGAGCCG		1018
OY	1169	cgctgaactcagttgtgccaaaaagtlitcaacacacacactlttggccaagcagcactgtg		1228
Dd	1017	CTCTGATGTGTGACCGCTCGGCGACGCCACACACTCTCTCTTCTGGGTAGCGGGGTGCACCAATG		1076
OY	1229	ccttggccaagcaacttggccgcgcgggaanaatcalcgtlcaacctloaagaatltgctlgacag		1288
Dd	1077	CTCTGGGCGACGAACCTGGCGCCGCTTCGAACTCGAAGTCATCTACCGTGTGTTCGACCG		1136
OY	1289	gattccctgaacttccacttgc	1310	
Dd	1137	CATTCCGACCTGCGCCTGGCC	1158	
<hr/>				
RESULT	8			
ID	011126			
AC	011126	standard: DNA; 1879 BP.		
DT	05-JUN-1991	(first entry)		
DE	Sequence encoding P450S1 and FeS-B.			
KM	Cytochrome P450; P450S1; P450S2; herbicide resistance; ds.			
OS	Streptomyces griseolus.			
FT	CDS			
FT	/+tag- a			
FT	misc_rna	128..1345		

FT	/*tag= b
FT	/product= P450S01
FT	1369. .1575
FT	misc_rna
FT	/tag= c
FT	/product= Fes-B protein
PN	W09103561-A.
PD	21-MAR-1991.
PF	27-AUG-1990; U04785.
PR	11-SEP-1989; US-405605.
PR	12-JAN-1990; US-464499.
PR	23-AUG-1990; US-569781.
PA	(DUPO ) DU PONT DE NEMOURS CO.
P1	Dean C, Harder PA, Leto KJ, Lichtner FT, Odell JT;
P1	O'Keefe DP, Omer CA, Romesser JA;
DR	WPI: 91-102077/14.
DR	P-PSDB: R11349, R11803.
PT	DNA encoding cytochrome P450 enzymes - and electron donating
PT	iron-sulphur proteins, used to confer herbicide resistance to
PT	plants and microorganisms
PS	Claim 13: page 151; 224pp; English.
CC	This sequence contains the coding information for (a) the cytochrome
CC	P450 enzyme, P450S01; and (b) the electron donor iron sulphur
CC	protein Fes-B. It is contained in a recombinant plasmid which is
CC	used to transform host Streptomyces species which in turn are used
CC	to coat a plant seed to transform the plant. The resultant trans-
CC	formants are resistant to herbicides.
CC	See also Q11127.
CC	Sequence 1879 BP; 301 A; 672 C; 645 G; 261 T;

[illegible]

```
QY 1057 locaccatccgattacagattcatggtcgctgcaactgaagaaggtgacagatcctg 1116
    || || || || || || || || || || || || || || || || || || || || ||
Db 1020 TCGCCACGCGGACATCGAGTCGAGGGGCACTCATCGCGGGCGGCGAGGGCGTGAACG 1079

QY 1117 tacgcagatgcgtctgctgctgctgctgctgctgctgctgctgctgctgctgct 1176
    || || || || || || || || || || || || || || || || || || || || ||
Db 1080 TCGTCACTCGATGACCCACCGGAGCGCACGCTGTACGAGGAGCCCGGACCCCTCGACA 1139

QY 1177 tcagtcgcagaaaggttccacacacccttggccacagcagcagcatctgtcttgcc 1236
    || || || || || || || || || || || || || || || || || || || || ||
Db 1140 TCCACCGCTCCGCGCGCCACCACTCGCTTCGGCTTGCGGCTGACCACTGCTGGGCC 1199

QY 1237 agcaccgtgcccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1275
    || || || || || || || || || || || || || || || || || || || || ||
Db 1200 AGAACCTCGCCCGCTGAGCTGAGGTGATCTCTCAACG 1238

RESULT 9
ID T58555 standard; cDNA; 1194 BP.
AC T58555:
DF 02-APR-1997 (first entry)
DE Streptomyces pristinaespiralis snbF gene.
KW Streptogramin B; antibiotic; biosynthesis; pristinamycin;
KM virginiamycin; piperolic acid; cyclodeamination; papa; snbA; snbF;
KW p1pA; 3-hydroxypicolinic acid; hydroxylation; ds.
OS Streptomyces pristinaespiralis.
FH Key Location/Qualifiers
FT mat_peptide 1..1194
FT /tag= a
FT /product= snbF
PD WO9601901-A1.
PN 25-JAN-1996.
PE 04-JUL-1995; F00889.
PR 08-JUL-1994; FR-008478.
PA (RHON) RHONE POULENC RORER SA.
PI Barriere JC, Blanc V, Blanchet F, Crouzet J;
PI Debussche L, Paris JM, Thibaut D, Bamas-Jacques N;
PI Dutruc-Rosset G, Farnachon A;
PI WPI: 96-097631/10.
DR P-PSDB; W11585.
FT New streptogramin B deriva. useful as antibiotics - produced by new
FT mutants of Streptomyces having altered genes for streptogramin B
FT biosynthesis
PS Claim 18: Page 113-114; 146pp: French.
CC The papa gene of S.pristinaespiralis is involved in the biosynthesis
CC of 4-dimethylamino-L-phenylalanine (DMPAPA), a precursor for
CC pristinamycin 1A. Upstream of the papa gene, on the complementary
CC strand, is the snbA gene coding for 3-hydroxypicolinic acid-AMP ligase.
CC The region between these two genes was sequenced and two open reading
CC frames were identified. The first (p1pA) decodes to an amino acid
CC sequence with homology to ornithine cyclodeaminase from Agrobacterium
CC tumefaciens. The p1pA gene product is likely to catalyse the cyclo-
CC demethylation of lysine, leading to production of piperolic acid.
CC Mutations in the p1pA gene were shown to affect piperolic acid
CC synthesis but not the synthesis of 3-hydroxypicolinic acid. The second
CC open reading frame (snbF) could be decoded to give a product with
CC homology to hydroxylases of the cytochrome P450 type.
CC Disruption of the p1pA and snbF genes can be used to produce
CC strains of S.pristinaespiralis which are unable to produce the
CC antibiotic pristinamycin I but which may be able to produce new,
CC modified forms of it.
SQ Sequence 1194 BP; 189 A; 522 C; 321 G; 162 T;
```

```
QY 794 gcagaagccgggaaccgacgcatcatgacatcgttgcacagcgacggtcaatggcgac 853
    || || || || || || || || || || || || || || || || || || || || ||
Db 603 CGCCCGCCCGCCGAGACGACATGATTCGCGACATCTGCCCGCCCGCTGGAGGGGAGAACG 662

QY 854 gataccagatgaagaagcagaagatgltgtgctgttactgtgtcggcgctgatatc 913
    || || || || || || || || || || || || || || || || || || || || ||
Db 663 ACTCACCGACGACGAGATGTCGTAATTCGGGGCGCTGCTGCTGATGGCCGACGCTTC 722

QY 914 ggtgtcaattctctcagatcagatgagatgattctctgtgcacaaagcccgagcatgc 973
    || || || || || || || || || || || || || || || || || || || || ||
Db 723 CACTTCATGCTGCTCGGCAACACCGTCTGTGCTGAAGAGACCAACCCCGGCGAGGC 782

QY 974 ggaagtcagtcagagcgtcccgagcgatattcaccgacgcttgcagagaactactccggc 1033
    || || || || || || || || || || || || || || || || || || || || ||
Db 783 CGCCCGCCCGCCCGACCGCTGATTCCTGATTCCTGCTGATTCGCTGATTCGCTGCTGCG 842

QY 1034 ctgcgtgtgtgcagatgagcgcacatcctcaccctcagatcagattcattggtgctgca 1093
    || || || || || || || || || || || || || || || || || || || || ||
Db 843 GCGCGGATCACCGTCATGAGCGCCCGGTCACCAACGACCGTCTGCGCGGACAC 902

QY 1094 gaagaaggtgacagatcctcgtacgcgacagatgctgtgtgctgtgagtcgcga 1133
    || || || || || || || || || || || || || || || || || || || || ||
Db 903 CATCCCGCCCGACGACGATGTCGTCCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 962

QY 1154 cgcctgcccagatgacgctgacatcagtcagtcgcaaaagttcacaacaccacttggca 1213
    || || || || || || || || || || || || || || || || || || || || ||
Db 963 CTTACCGACCCCGACACCACTTCGACCTGCGCCCGGAGAGCGCCGACGATCGCTTGGCGCA 1022

QY 1214 cgcagaccatctgtgcttgcgcagacatctgcccgcgcggaatcaltcgtacacctca 1273
    || || || || || || || || || || || || || || || || || || || || ||
Db 1023 CGCATCCACTACTGCTCGGGCGCCCGCTGCGCGCTGAGGGCGCATTCGCGCTGGA 1082

QY 1274 ggaatgctgacccagatctcctgacttccatctgacccggtgcccagatcagaca 1333
    || || || || || || || || || || || || || || || || || || || || ||
Db 1083 AGCCCTTTCGACCGATTCGCCGACTTCTGCCCAACGACGCGCAAACTGCGCTACA 1142

QY 1334 gacgcagctgcagcggtgcagcgacatccctct 1370
    || || || || || || || || || || || || || || || || || || || || ||
Db 1143 CCGCAGCGACTGTTGGCGCTCAACGACCTGCGCTG 1179

RESULT 10
ID T58553
AC T58553 standard; cDNA; 4496 BP.
DF 02-APR-1997 (first entry)
DE Streptomyces pristinaespiralis snbA and papa intergenic region.
KW Streptogramin B; antibiotic; biosynthesis; pristinamycin;
KM virginiamycin; piperolic acid; cyclodeamination; papa; snbA; snbF;
OS Streptomyces pristinaespiralis.
FH Key Location/Qualifiers
FT cds complement (1..332)
FT /tag= a
FT /product= snbA
FT /note= "N-terminal coding region only, i.e. a
FT partial open reading frame"
FT 607..1674
FT cds
FT /tag= b
FT /product= p1pA
FT 1800..2996
FT /tag= c
FT /product= snbF
FT 3018..4496
FT /tag= d
FT /product= papa
FT /note= "N-terminal coding region only, i.e. a
FT partial open reading frame"
```

PA (RHON ) RHONE POULENC RORER SA.  
PI Barriere JC, Blanc V, Blanche F, Crouzet J;  
PI Debussche L, Paris JM, Thibaut D, Bamas-Jacques N;  
PI Dutric-Rosset G, Famechon A;  
DR WPI: 96-097631/10.  
P-PSDB: W11584, W11585.  
PT New streptogramin B derivs. useful as antibiotics - produced by new  
PT mutants of Streptomyces having altered genes for streptogramin B  
PT biosynthesis  
PS Example 1: Page 107-111: 146pp: French.  
CC The papa gene of *S. pristinaespiralis* is involved in the biosynthesis  
CC of 4-dimethylamino-L-phenylalanine (DMAPA), a precursor for  
CC pristinamycin 1A. Upstream of the papa gene, on the complementary  
CC strand, is the snbA gene coding for 3-hydroxypicolinic acid-AMP ligase.  
CC The region between these two genes was sequenced and two open reading  
CC frames were identified. The first (papa) decodes to an amino acid  
CC sequence with homology to ornithine cyclodiaminase from *Agrobacterium*  
CC *lumea*clens. The papa gene product is likely to catalyse the cyclo-  
CC deamination of lysine, leading to production of picolinic acid.  
CC Mutations in the papa gene were shown to affect picolinic acid  
CC synthesis but not the synthesis of 3-hydroxypicolinic acid. The second  
CC open reading frame (snbA) could be decoded to give a product with  
CC homology to hydroxylases of the cytochrome P450 type.  
CC Disruption of the papa and snbA genes can be used to produce  
CC strains of *S. pristinaespiralis* which are unable to produce the  
CC antibiotic pristinamycin I but which may be able to produce new,  
CC modified forms of it.  
SQ Sequence 4496 BP: 659 A; 1980 C; 1321 G; 536 T;

Query Match 5.9%; Score 83.4; DB 1; Length 4496;  
Best Local Similarity 45.7%; Pred. No. 8.8e-13;  
Matches 291; Conservative 0; Mismatches 346; Indels 0; Gaps 0;

OY 734 ctgcgagagggccaagagcgctctacgactatctatgataccgatactgagcaagcag 793  
DB 2342 CGTCAAGAAACCCATGGCGCCCATGACGCTCACTCCACGACACGACCGCGCGG 2401  
OY 794 gcagaaaggcgaagccagcgtatcagacgtgtgccaagcgccaggtcaatggcgacc 853  
DB 2402 CGCGCGCGCGCGGACGACGCTGATCTCCGACATCTGCGCGCGCGCGCTGAGGCGGACG 2461  
OY 854 gatcaccagtgcagaaagccaagagatgtgtgcctgtactgtgcgcgcgcgtgatac 913  
DB 2462 ACTACCGCAGACAGATCTCGCAATTCGGGCGCTGCTGATGACGCGCACGCTCTC 2521  
OY 914 ggtgtcaattctcagctcagctgagctgtcctcgtgccaagaagccgagatgcga 973  
DB 2522 CACCTCATCTGCTGCGCAACCGTGTGCTGTGAAGACCCCGGGCGGAGGC 2581  
OY 974 ggaagctgatacgagcgtcccgagcgtatccagcgcgtctgaggaatactaccgcgctt 1033  
DB 2582 CGCGCGCGCGCGGACGCGCTGCTGATCTCCGCGCGCGCTGATGAGAACTACTGCGGCTGCG 2641  
OY 1034 ctgcgtgtgttcgagatggcgcgcatcctacccgattacgaatttcaatgagcgttgcaact 1093  
DB 2642 GCGCGCGGATCACCGTCAATGCGCGCGCTGACACCAAGAGACCGCTCTCGCGGACACAC 2701  
OY 1094 gaagaaagtgagcaagatcctgtacccgagatgctgtcgtcgtgattgagcggaataa 1153  
DB 2702 CATCCCCCGGAGCGATGGTGTGCGCTCTCCCTCTGCGCAACACACAGCAACAGGT 2761  
OY 1154 cgcctgcgcagatgcagctgcaacttcaatgcgcaaaagtttccacacccaccttggcca 1213  
DB 2762 CTTCACCGAGACCGGACGACCTCGACCTCGCGCGGAGAGCGCGCGCGCTTCGCGGCA 2821  
OY 1214 cggcagccatctgtgcttggccaagacgtgcccgcgggaaatacatgcaacctcaaa 1273  
DB 2822 CGGCAATCACTACTGCTGTGGCGCGCGCTGCGCGCTGAGAGGCGCGCATCCCTTGGA 2881  
OY 1274 ggaatggtcgcagcagatctcgaattctcatatgcccgcgggtgcccagattcagcaaa 1333  
DB 2882 AGCCCTTTCGACGATTCGCCGACTTCTGCGCCACCGACGCGGCAAAACTGCGCTACCA 2941

OY 1334 ggcggcagatcgtcagcgagcgtgcagcagcactcctctg 1370  
DB 2942 CCGCGACGAGACTTTCGCGCTCAAGAACTGCGCGCTG 2978

## RESULT 11

ID V21187 standard; DNA; 53789 BP.

AC V21187;

DT 24-JUL-1998 (first entry)

DE Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.

KW Amycolatopsis mediterranei: rifamycin; synthesis; gene cluster;

OS Amycolatopsis mediterranei.

FH Key location/Qualifiers

FT CDS 1825..15543

FT /tag= a

FT /label= ORF\_A

FT /product= "polyketide synthase"

FT CDS 15550..30759

FT /tag= b

FT /label= ORF\_B

FT /product= "polyketide synthase"

FT CDS 30895..36060

FT /tag= c

FT /label= ORF\_C

FT /product= "polyketide synthase"

FT CDS 36259..41325

FT /tag= d

FT /label= ORF\_D

FT /product= "polyketide synthase"

FT CDS 41373..51614

FT /tag= e

FT /label= ORF\_E

FT /product= "polyketide synthase"

FT CDS 51713..5293

FT /tag= f

FT /label= ORF\_F

FT /product= "polyketide synthase"

PN MO9807868-A1.

PD 26-FEB-1998.

PF 18-AUG-1997; E04495.

PR 20-AUG-1996; ED-810551.

PA (NOVS ) NOVARTIS AG.

PI Engel N, Schupp T, Toupet C;

PT WPI: 98-169172/15.

P-PSDB: W52845-W52850.

PT Amycolatopsis mediterranei rifamycin synthesis gene cluster - used

PT to produce rifamycin and rifamycin analogues

PS Claim 4; Page 53-102; 205pp: English.

CC The present sequence represents a Amycolatopsis mediterranei rifamycin

CC synthesis gene cluster DNA fragment from the present invention. The

CC DNA fragment comprises a DNA region involved directly or indirectly

CC in the gene cluster responsible for rifamycin synthesis, including

CC the adjacent DNA regions to the right and left which, by reason of

CC their function in connection with rifamycin biosynthesis, qualify

CC as constituents of this rifamycin gene cluster, and functional

CC fragments, derivatives or constituents of these. The Amycolatopsis

CC mediterranei rifamycin synthesis gene cluster DNA fragment can be used

CC for producing rifamycin, rifamycin analogues or precursors. It can also

CC be used for inactivating or modifying genes involved in ansamycin or

CC rifamycin biosynthesis. The DNA can be used for constructing mutant

CC actinomycetes strains from which the natural rifamycin or ansamycin

CC biosynthesis gene cluster has been partly or completely deleted. The

CC DNA fragment can be used for assembling a library of polyketide

CC syntheses, which can be used for assembling a library of polyketides.

CC A hybridisation probe of the invention can be used for identifying DNA

CC fragments involved in the biosynthesis of ansamycins.

SO Sequence 53789 BP: 6707 A; 19183 C; 20504 G; 7395 T;

Query Match 5.5%; Score 77.8; DB 1; Length 53789;





```

Db 1161 GAGCGGAGCTCGGCTCGACCGGTGCCCGCCAAACCGGGACCCAGCTTTCCCAACGC 1220
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 1164 atgcacgtcgacattcaagtcgcaaaagtttcacacacacatttggcagcgagccat 1223
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1221 GACCGGATCGACGTCGACCCCTACGCCCAATCAGCAATCTCGGTTGCGGACACGCGCTCCAC 1280
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 1224 ctggccttggcagcacacccctggccgcccgggaatatatgtcaacctcaagaatgct 1282
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1281 CACTGCTCGGCGCTCTCTGCGCCCGGATGCACTCAGAGTGACCTGAGGTCCTGCT 1339
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
ID Q11127
AC Q11127 standard; DNA; 1998 BP.
DT 05-JUN-1991 (first entry)
DE Sequence encoding P450SU2 and FeS-A.
KW P4Cytochrome P450; P450SU1; P450SU2; herbicide resistance; ds.
OS Streptomyces griseolus.
FH Key
FT misc_rna
FT 195..1403
FT /tag= a
FT /product= P450SU2
FT misc_rna
FT 1452..1643
FT /tag= b
FT /product= FeS-A protein

W09103561-A.
PD 21-MAR-1991.
PF 27-AUG-1990; U04785.
PF 11-SEP-1989; US-405605.
PR 12-JAN-1990; US-464499.
PR 23-AUG-1990; US-569781.
PA (DUPO.) DU PONT DE NEMOURS CO.
PI Dean C, Harder PA, Leto KJ, Lichtner FT, Odell JT;
PI O'Keefe DP, Omer CA, Komesser JA;
PI WPI: 91-102077/14.
DR P-PSDB; R11350, R11804.
PT DNA encoding cytochrome P450 enzymes - and electron donating
PT iron-sulphur proteins, used to confer herbicide resistance to
PT plants and microorganisms
PT
PS Claim 15; page 158; 224bp; English.
CC This sequence contains the coding information for (a) the cytochrome
CC P450 enzyme, P450SU2; and (b) the electron donor iron sulphur
CC protein FeS-A. It is contained in a recombinant plasmid which is
CC used to transform host streptomyces species which in turn are used
CC to coat a plant seed to transform the plant. The resultant trans-
CC formants are resistant to herbicides.
CC See also Q11126.
SO Sequence 1998 BP; 272 A; 839 C; 620 G; 267 T;

```

[illegible][illegible]

```

RESULT 14
ID 045569
AC 045569;
DT 04-DEC-1993 (first entry)
DE Sequence soyC and soyC genes encoding cytochrome P450soy and
DE ferredoxin-soy.
KM P450soy; soyC gene; ferredoxin-soy; soyB gene; oxidation;
KM haem protein; ss.
OS Streptomyces griseus ATCC 13273.
FH Key location/Qualifiers
FT cds
FT 149..1384
FT /*tag= a
FT /product= P450-soy protein
PN W09J12236-A.
PD 24-JUN-1993.
PE 16-DEC-1992; U10885.
PR 16-DEC-1991; US-807001.
PA (DUPO) DU PONT DE NEMOURS & CO E. I.
PI Omer CA, Sariaslani FS, Trower MK;
DR WPI: 93-214178/26.
DR P-PSDB: R38309.
PT Constitutive expression of P450 SOY (SOYC) and ferredoxin soy
PT (soyB) in Streptomyces - used for oxidn. of organic chemicals
PT Example: Figure 2: 45pp; English.

```



```

Qy 1069 attacgaatttcattgagcgtgcaactgagaagaagtgaccagatcctgtactaaccgagatgc 1128
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1050 ACCTGGAGATCGGGGTGTGCGCATCCCGCAGTACAGCAGCGTGTGTGCGGACGCG 1109
Qy 1129 tctctgacctgagatgagcgcaaaacgacctgcccagatgcaagtcgaactcaagtcgcaaa 1188
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1110 CGGCCAACCCGACCGAGAGCAGTTCGCCGACCCCGACCGCTCCGACGTCACCCGCGACA 1169
Qy 1189 aggtttcacacacacaccttggcagcgcaagcaccatctgtaccttgccagcaacctgccc 1248
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1170 CCCCGCGCCACCTGTCTGCGGAGGCAACCACTTCTGATGGGCGCGCGCTGGCCA 1229
Qy 1249 gccgggaatcattcgtcaccctcaaggaaatgctgaccagatctcgaacttccattg 1308
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1230 AGCTGGAGGGGCGAGGTGGCGCTGTGCGGCGCTTCCCGCTCTGTGCGCTGG 1289
Qy 1309 ccccggtgcacca 1321
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1290 GAATCGACGCCGA 1302

```

Search completed: October 4, 2000, 06:40:28  
 Job time: 12237 sec



Db	241	tttccagagcagtgccggttcattccctccgcggaagccggcgaaagcctacgattcatccc	300
QY	459	acctcgagagatcccgcccgagacagcgcagttctcgtgcgtgccaacaagtgtgtgc	518
Db	301	acctcgagtgatcccgcccgagcagcgcaattctcgtcgtgccaacaagtgtgtgc	360
QY	519	atgcgcgtgtgtgataagctgcggaacaacggatccagagatgtgcgtctgcgtgacgag	578
Db	361	atgcgcggtgtgtgataagctgcggaacaacggatccagagatgtgcgtctgcgtgacgag	420
QY	579	agccctgcgcccgcgaagagacagtgcgaacttaccgcgagagatcacgcgcgaacctcccgata	638
Db	421	agccctgcgcccgcgaagagacagtgcgaacttaccgcgagagatcacgcgcgaacctcccgata	480
QY	639	cgcatctctcatgtcgtctcgcgaagctctaccggaagaagatactccgcacitgaaataccta	698
Db	481	cgcatctctcatgtcgtctcgcgaagctctaccggaagaagatactccgcacitgaaataccta	540
QY	699	acggatacgaatgacccgtcccgatgagagatatacctctgcgaaggccaagagagcgctc	758
Db	541	acggatacgaatgacccgtcccgatgagagatatacctctgcgaaggccaagagagcgctc	600
QY	759	tacgactatctgtatataccgatalcattcgacaacgcgacgacgaagccgggaaaccgagctac	818
Db	601	tacgactatctgtatataccgatalcattcgacaacgcgacgacgaagccgggaaaccgagctac	660
QY	819	agctcgtctgcgaacgcgcgcaggttcaatagggcgagaccgataccacgagtacgaaagccaaagag	878
Db	661	agctcgtctgcgaacgcgcgcaggttcaatagggcgagaccgataccacgagtacgaaagccaaagag	720
QY	879	atgctgtgcgcgttactacgctgcgcgcgtctgatacgaatgtgtaattctccacgattcacg	938
Db	721	atgctgtgcgcgttactacgctgcgcgcgtctgatacgaatgtgtaattctccacgattcacg	780
QY	939	atggagatctcctgcgcaaaaagcccgagacatcgcgacgaagctgtatctcgacgctcccgacgt	998
Db	781	atggagatctcctgcgcaaaaagcccgagacatcgcgacgaagctgtatctcgacgctcccgacgt	840
QY	999	attccacagcgccttgcgaggaactactctccgcgccttctcgtctgtgttcgcgatgtgcgcatc	1058
QY	1059	ctcacccctcgatitacgagttctcatagcgtgttcaactgaagaagaagtgaaccagatccctgcata	1118
Db	901	ctcacccctcgatitacgagttctcatagcgtgttcaactgaagaagaagtgaaccagatccctgcata	960
QY	1119	ccgagagatgctgtctctgcgccttgcgatagcgcgcaaaacgcgtcccggaatgcacgtctgcattc	1178
Db	961	ccgagagatgctgtctctgcgccttgcgatagcgcgcaaaacgcgtcccggaatgcacgtctgcattc	1020
QY	1179	agttcgcaaaaaggttttaacacacacaccttttgcgcacgcgacgacatctgtgctcttgcgag	1238
Db	1021	agttcgcaaaaaggttttaacacacacaccttttgcgcacgcgacgacatctgtgctcttgcgag	1080
QY	1239	caactgcgcccgcggaagaatcatctgtacacctcaagaagtgtgtgacgaagatctcttcgac	1299
Db	1081	caactgcgcccgcggaagaatcatctgtacacctcaagaagtgtgtgacgaagatctcttcgac	1140
QY	1299	ttctccatctgcgcccggtgtcccgagatctcaagacacaaagcggcatctgtctcagcggctgtcag	1359
Db	1141	ttctccatctgcgcccggtgtcccgagatctcaagacacaaagcggcatctgtctcagcggctgtcag	1200
QY	1359	gcaatccctctgtctctgtgatccggcgacatccaaagcggata 1400	
Db	1201	gcaatccctctgtctctgtgatccggcgacatccaaagcggata 1242	

```

RESULT      2
US-09-068-132-2
; Sequence 2, Application US/09068132
; Patent No. 6117661
; GENERAL INFORMATION:
; APPLICANT: Wong, Yuet-lok

```

```

1  APPLICANT: Flitsch, Sabine Iahja
2  APPLICANT: Nickerson, Darren Paul
3  APPLICANT: Hart, Alwyn James
4  TITLE OF INVENTION: Mutant Mono-Oxygenase Cytochrome P450C4M
5  FILE REFERENCE:
6  CURRENT APPLICATION NUMBER: US/09/068,132
7  CURRENT FILING DATE: 1998-09-14
8  EARLIER APPLICATION NUMBER: GB 95 22407.7
9  EARLIER FILING DATE: 1995-11-01
10 NUMBER OF SEQ ID NOS: 2
11 SOFTWARE: Microsoft Word Office 98 for Macintosh
12 SEQ ID NO 2
13     LENGTH: 1242
14     TYPE: DNA
15 ORGANISM: Artificial Sequence
16 FEATURE:
17 NAME/KEY: P450cam C334A
18 LOCATION: 1..1242
19 OTHER INFORMATION: Mutant with Cys-334 (Tgc at nucleotides 1000
20 US-09-068-132-2
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991

```

Query Match	88.4%	Score 1238.8	DB 5	Length 1242
Best Local Similarity	99.8%	Pred. NO. 0		
Matches 1240	Conservative 0	Mismatches 2	Indels 0	Gaps
QY	159	acgactgaaacacatacaagaacagccacatctctgcccctctgcccaccccatctgcagaag	218	
Db	1	acgactgaaacacatacaagaacagccacatctctgcccctctgcccaccccatctgcagaag	60	
QY	219	caaccctgtattcgtgacttcgacatgtacaaatccgttgaaatctgtctgcgcggtgcagaag	278	
Db	61	caaccctgtattcgtgacttcgacatgtacaaatccgttgaaatctgtctgcgcggtgcagaag	120	
QY	279	gctctgagcattctcgcagaagaatcaaacgtaccagatctcgtgtgtgaaatccgtgcagaagc	338	
Db	121	gctctgagcattctcgcagaagaatcaaacgtaccagatctcgtgtgtgaaatccgtgcagaagc	180	
QY	339	ggacaactgatactgcacactcgcgcacactgtacatcgtctgagccctatgaaagatattaccgcac	398	
Db	181	ggacaactgatactgcacactcgcgcacactgtacatcgtctgagccctatgaaagatattaccgcac	240	
QY	399	ttttccagcgagctgcccgttcatccctcgtgaaagccgcggaagcctaagacttattccc	458	
Db	241	ttttccagcgagctgcccgttcatccctcgtgaaagccgcggaagcctaagacttattccc	300	
QY	459	acctcgatggaatccgcgccgcgaagcagcgccaagtctctgtcgtctgcacacaagaattgttgc	518	
Db	301	acctcgatggaatccgcgccgcgaagcagcgccaagtctctgtcgtctgcacacaagaattgttgc	360	
QY	519	atgcgcgtgtgtgataagcttgagaaaccggaatccagaagcttgacctgtctgcgtatcgaag	578	
Db	361	atgcgcgtgtgtgataagcttgagaaaccggaatccagaagcttgacctgtctgcgtatcgaag	420	
QY	579	agccgcgcgcgcgcgaaggaagatgcaatttcaaccgagagctacgcgaaacccctcccgata	638	
Db	421	agccgcgcgcgcgcgaaggaagatgcaatttcaaccgagagctacgcgaaacccctcccgata	480	
QY	639	cgcatcttcattctgtctcgcagagcttaccgcgaagaagatatcccgcatcttgaataacta	698	
Db	481	cgcatcttcattctgtctcgcagagcttaccgcgaagaagatatcccgcatcttgaataacta	540	
QY	699	acggatcagaatgaccccgctccggatgtgcagcatgatccttcgcgaagcccaagagcgctc	758	
Db	541	acggatcagaatgaccccgctccggatgtgcagcatgatccttcgcgaagcccaagagcgctc	600	
QY	759	tacgcgactctgtataccgatactatgaaagaacgcgaagccgcggaacgcgaagcgtctc	818	
Db	601	tacgcgactctgtataccgatactatgaaagaacgcgaagccgcggaacgcgaagcgtctc	660	
QY	819	agcatcgtctgcacaagcgccaggttcaatgtgagccacgcgataccacgtgaagaagccagaag	878	

Db	661	agcatcgtctgcccaacgcccaggtcaatagggcgacgcgatccacccagtgacgaagccaaagag	720
Qy	879	atgtgtgacgccttactatgtgttcgcgcgcgcctggatatacagttgtgtcaattctctcaagcttcaag	938
Db	721	atgtgtgacgccttactatgtgttcgcgcgcgcctggatatacagttgtgtcaattctctcaagcttcaag	780
Qy	939	atgtgaattccctggcgcacaaaagcccggagacatccgcgaagaagctgtatctgaagcgtcccgagcgt	998
Db	781	atgtgaattccctggcgcacaaaagcccggagacatccgcgaagaagctgtatctgaagcgtcccgagcgt	840
Qy	999	attccagacgccttgcgcgaagactactccgcgcgccttctcgtctgtatcttgcgatatgcgcgcatc	1051
Db	841	attccagacgccttgcgcgaagactactccgcgcgccttctcgtctgtatcttgcgatatgcgcgcatc	900
Qy	1059	ctcacctccgcattcgaatcgaagttcatatgtgcgcgtgcacactgaagaaggtgtacccaagatccgtcta	1111
Db	901	ctcacctccgcattcgaatcgaagttcatatgtgcgcgtgcacactgaagaaggtgtacccaagatccgtcta	960
Qy	1119	ccgcagagatgcgtctcgtgcgcctggatagacgcgcgaacacgcctcgcgcgccttgcacgtccgacttc	1177
Db	961	ccgcagagatgcgtctcgtgcgcctggatagacgcgcgaacacgcctcgcgcgccttgcacgtccgacttc	1021
Qy	1179	agtcgcgcacaaaggtttcaacacacacacacacttgcgcacgcgcgcacacatctgttgccttgcgcag	1231
Db	1021	agtcgcgcacaaaggtttcaacacacacacacacttgcgcacgcgcgcacacatctgttgccttgcgcag	1081
Qy	1239	cacctctgcgcgcgcgcgcgaataatcatctgtacaccttaagaagaatgctgtgcgcgcgcgcgcgcgcgc	1291
Db	1081	cacctctgcgcgcgcgcgcgaataatcatctgtacaccttaagaagaatgctgtgcgcgcgcgcgcgcgcgc	1141
Qy	1299	tctctccattgc	1351
Db	1141	tctctccattgc	1201
Qy	1359	gcactccctcgtcgtctgtgagttccgc	1400
Db	1201	gcactccctcgtcgtctgtgagttccgc	1242
RESULT 3			
5212296-16			
Patent No. 5212296			
APPLICANT: DEAN, CAROLINE; HARDER, PATRICIA A.; LEITO, KENNETH			
J.J.; O'KEEFE, DANIEL P.; OMER, CHARLES A.; ROMESSER, JAMES A.			
TEPPERMAN, JAMES M.			
TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING			
CYTOCHROMES			
NUMBER OF SEQUENCES: 19			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: US/07/569, 781			
FILING DATE: 23-AUG-1990			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: 464, 499			
FILING DATE: 12-JAN-1990			
APPLICATION NUMBER: 405, 605			
FILING DATE: 11-SEP-1989			
SEQ ID NO: 16:			
LENGTH: 1221			
5212296-16			

Query Match	6.38	Score 88.2	DB 7	Length 1221
Best Local Similarity	46.08	Pred. No. 3.3e-14		
Matches 377; Conservative	0	Mismatches 433	Indels 9	Gaps 2

[illegible][illegible]

```

RESULT 4
5212296-5
; Patent NO. 5212296
; APPLICANT: DEAN, CAROLINE; HARDER, PATRICIA A.; LETO, KENNETH
; J.; O'KEEFE, DANIEL P.; OMER, CHARLES A.; ROMESSER, JAMES A.
; TEPPERMAN, JAMES M.
; TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING
; CYTOCHROMES
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/569,781
; FILING DATE: 23-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 464,499
; FILING DATE: 12-JAN-1990
; APPLICATION NUMBER: 405,605
; FILING DATE: 11-SEP-1989
; SEQ ID NO:5
; LENGTH: 1879
5212296-5

```

Query Match	6.3%;	Score 88.2;	DB 7;	Length 1879;
Best Local Similarity	46.0%;	Pred. No. 3.8e-14;		







```
REGISTRATION NUMBER: 32,692
REFERENCE/DOCKET NUMBER: CR-9000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-892-7342
TELEFAX: 302-892-7949
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1735 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-102-863-10
```

```
Query Match 5.2% Score 73.6; DB 1; Length 1735;
Best Local Similarity 47.5% Pred. No. 2.4e-10;
Matches 251; Conservative 0; Mismatches 274; Indels 3; Gaps 1;
```

```
QY 743 ggcgaagagcgctctacgactatctatccgatacgcgaacgcaagcagaagcc 802
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 777 ggcggcgacgacgactcgagagctgacggcgccgctgattcgacccgagagggcgagcc 836
QY 803 gggaaacgacgctacgacatcgltgccaacggcgcaagltcaatggcgacgataccag 862
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 837 gggtagcggcctccctgagcagactgattccaccgggagccgagcgagcgctgacgcg 896
QY 863 tgaagaaagcaagagatgltggtctgttctgtggtggcgctgagatggtgtcaaa 922
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 897 cgaaacgctgctgcttgcgctgctgctgctgctgctgctgctgctgctgctgctgctg 956
QY 923 ttctcagcttcaagatgagtlcttgccaagaagccggagacatcgccaagagatgat 982
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 957 catgattcgcgtcgacgacgttccacgctgctgaccccgacacgacgagcgcgctgacg 1016
QY 983 cgaagctcccgagcgatltccagccgcttgcgaggaactactcgagcgcttctcgctgt 1042
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1017 ggcggcgagcagacacgacgctgctgctgctgctgctgctgctgctgctgctgctgct 1076
QY 1043 tgcgagtgccgacatcctacgtccgattac--gagtttcaatgctgctgcaactgaa 1099
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1077 cgagggccttccagcgcctgagcagcagcagcagcagcagcagcagcagcagcagcag 1136
QY 1100 aggtgacacagatcctgctacgacgacatgctgtctgctgagtgagcgcgaaagcgctg 1159
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1137 ggggagggcggtgcttctgacactgctgacacgacgacgacgacgacgacgacgacgac 1196
QY 1160 ccgagatcgactcgactcagtcgcaaaaggttccacacacacacacacacacacacacac 1219
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1197 ggcgagacgactcgactgagacgacgacgacgacgacgacgacgacgacgacgacgacg 1256
QY 1220 ccaatcgcttgccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1267
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1257 ccaccagtgcttgggcgacgacgacgacgacgacgacgacgacgacgacgacgacgacg 1304
```

```
RESULT 8
PCT-US92-10885-10
; Sequence 10, Application PC/TUS9210885
```

```
; GENERAL INFORMATION:
; APPLICANT: SARIASLANI, SIMA
; TITLE OF INVENTION: CONSTITUTIVE
; TITLE OF INVENTION: EXPRESSION OF P450SOY
; TITLE OF INVENTION: AND FERREDOXIN-SOY IN
; TITLE OF INVENTION: STREPTOMYCES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: USA
```

```
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 Inch,
MEDIUM TYPE: 1.0 MB
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh System, 6.0
SOFTWARE: Microsoft Word, 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/10885
FILING DATE: 19921216
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: GALLEGOS, R. THOMAS
REGISTRATION NUMBER: 32,692
REFERENCE/DOCKET NUMBER: CR-9000-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-892-7342
TELEFAX: 302-892-7949
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1735 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PCT-US92-10885-10
```

```
Query Match 5.2% Score 73.6; DB 6; Length 1735;
Best Local Similarity 47.5% Pred. No. 2.4e-10;
Matches 251; Conservative 0; Mismatches 274; Indels 3; Gaps 1;
```

```
QY 743 ggcgaagagcgctctacgactatctatccgatacgcgaacgcaagcagaagcc 802
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 777 ggcggcgacgacgactcgagagctgacggcgccgctgattcgacccgagagggcgagcc 836
QY 803 gggaaacgacgctacgacatcgltgccaacggcgcaagltcaatggcgacgataccag 862
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 837 gggtagcggcctccctgagcagactgattccaccgggagccgagcgagcgctgacgcg 896
QY 863 tgaagaaagcaagagatgltggtctgttctgtggtggcgctgagatggtgtcaaa 922
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 897 cgaaacgctgctgcttgcgctgctgctgctgctgctgctgctgctgctgctgctgctg 956
QY 923 ttctcagcttcaagatgagtlcttgccaagaagccggagacatcgccaagagatgat 982
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 957 catgattcgcgtcgacgacgttccacgctgctgaccccgacacgactggcgcgctgacg 1016
QY 983 cgaagctcccgagcgatltccagccgcttgcgaggaactactcgcgcttctcgctgt 1042
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1017 ggcggcgagcagacacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 1076
QY 1043 tgcgagtgccgacatcctacgtccgattac--gagtttcaatgctgctgcaactgaa 1099
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1077 cgagggccttccagcgcctgagcagcagcagcagcagcagcagcagcagcagcagcag 1136
QY 1100 aggtgacacagatcctgctacgacgacatgctgtctgctgagtgagcgcgaaagcgctg 1159
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1137 ggggagggcggtgcttctgacactgctgacacgacgacgacgacgacgacgacgacgac 1196
QY 1160 ccgagatcgactcgactcagtcgcaaaaggttccacacacacacacacacacacacacac 1219
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1197 ggcgagacgactcgactgagacgacgacgacgacgacgacgacgacgacgacgacgacg 1256
QY 1220 ccaatcgcttgccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1267
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1257 ccaccagtgcttgggcgacgacgacgacgacgacgacgacgacgacgacgacgacgacg 1304
```

```
RESULT 9
5212296-17
; Patent No. 5212296
; APPLICANT: DEAN, CAROLINE; HARDER, PATRICIA A.; LETO, KENNETH
```





```

1 STREET: Two Milltia Drive
2 City: Lexington
3
4 STATE: MA
5
6 COUNTRY: USA
7
8 ZIP: 02173
9
10 COMPUTER READABLE FORM:
11 MEDIUM TYPE: Floppy disk
12
13 COMPUTER: IBM PC compatible
14
15 OPERATING SYSTEM: PC-DOS/MS-DOS
16
17 SOFTWARE: Patentln Release #1.0, Version #1.25
18
19 CURRENT APPLICATION DATA:
20
21 APPLICATION NUMBER: US/08/444,623
22
23 FILING DATE:
24
25 CLASSIFICATION: 435
26
27 PRIOR APPLICATION DATA:
28
29 APPLICATION NUMBER: 08/095,734
30
31 FILING DATE: 22-JUL-1993
32
33 APPLICATION NUMBER: 07/711,334
34
35 FILING DATE: 06-JUN-1991
36
37 APPLICATION NUMBER: 07/367,894
38
39 FILING DATE: 19-JUN-1989
40
41 APPLICATION NUMBER: PCT/US90/03451
42
43 FILING DATE: 18-JUN-1990
44
45 APPLICATION NUMBER: PCT/US89/02962
46
47 FILING DATE: 07-JUL-1989
48
49 APPLICATION NUMBER: 07/361,944
50
51 FILING DATE: 05-JUN-1989
52
53 APPLICATION NUMBER: 07/223,089
54
55 FILING DATE: 22-JULY-1988
56
57 APPLICATION NUMBER: 07/216,390
58
59 FILING DATE: 07-JUL-1988
60
61 APPLICATION NUMBER: 07/163,546
62
63 FILING DATE: 03-MAR-1988
64
65 APPLICATION NUMBER: PCT/US88/00614
66
67 FILING DATE: 29-FEB-1988
68
69 APPLICATION NUMBER: 07/020,451
70
71 FILING DATE: 02-MAR-1987
72
73 ATTORNEY/AGENT INFORMATION:
74
75 NAME: Granahan, Patricia
76
77 REGISTRATION NUMBER: 32,227
78
79 REFERENCE/DOCKET NUMBER: WI93-11M2
80
81 TELECOMMUNICATION INFORMATION:
82
83 TELEPHONE: 617-861-6240
84
85 TELEFAX: 617-861-9540
86
87 INFORMATION FOR SEQ ID NO: 1:
88
89 SEQUENCE CHARACTERISTICS:
90
91 LENGTH: 4394 base pairs
92
93 type: nucleic acid
94
95 STRANDEDNESS: double
96
97 TOPOLOGY: linear
98
99 MOLECULE TYPE: DNA (genomic)
100
101 OS-08-444-623-1

```

Query Match	2.6%;	Score 37;	DB 3;	Length 4394;
Best Local Similarity	50.3%;	Pred. No. 1.2;		
Matches	91;	Conservative	0;	Mismatches 90; Indels 0; Gaps 0;
QY 1220	ccatctgtgccttggccagcaaccttggccgccgggaaatcatctgcacccctcaagaaatg			1279
Db 2066	CTATTGGGCTTCCGGTTCGCTCCGCGCCCTGCTAAGAGTCGCGCCGCCACAGCGCCGAGG			2125
QY 1280	gctaacccaggaattccttgactctccatctggcccccgggtgcccaagtatccacaagaagcgg			1339
Db 2126	GGTGTTCTGCTGCGCGGCGCACCTCCAAATCCGAGGGGTGGCGGTGCGMAAATGCCGCCGC			2185
QY 1340	catcttcagcggcgtgtgcagcgcactccctcttggtcttcgggatacccgagactaccaaaagcgt			1399
Db 2186	CGACGCGCGCAGCGTGGGCCCATGTTGTTGTGTGACCAAGGTGGGGGGCGCCAAAGAGGGCGC			2245
QY 1400	a 1400			
Db 2246	A 2246			

RESULT 14  
US-08-471-869-1  
Sequence 1, Application US/08471869  
Patent No. 6022745  
GENERAL INFORMATION:  
APPLICANT: Aldovini, Anna  
APPLICANT: Young, Richard A.  
TITLE OF INVENTION: Homologously Recombinant Slow Growing  
TITLE OF INVENTION: Mycobacteria and Uses Therefor  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Millitia Drive  
CITY: Lexington  
STATE: MA  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/471,869  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/08267  
FILING DATE: 22-JUL-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/095,734  
FILING DATE: 22-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/711,334  
FILING DATE: 06-JUN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/367,894  
FILING DATE: 19-JUN-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US90/03451  
FILING DATE: 18-JUN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US89/02962  
FILING DATE: 07-JUL-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/361,944  
FILING DATE: 05-JUN-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/223,089  
FILING DATE: 22-JUL-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/216,390  
FILING DATE: 07-JUL-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/163,546  
FILING DATE: 03-MAR-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US88/00614  
FILING DATE: 29-FEB-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/020,451  
FILING DATE: 02-MAR-1987  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: WH193-11MA2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-861-6240  
TELEFAX: 617-861-9540  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:

RESULT 14  
US-08-471-869-1  
Sequence 1, Application US/08471869  
Patent No. 6022745  
GENERAL INFORMATION:  
APPLICANT: Aldovini, Anna  
APPLICANT: Young, Richard A.  
TITLE OF INVENTION: Homologously Recombinant Slow Growing  
TITLE OF INVENTION: Mycobacteria and Uses Therefor  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Millitia Drive  
CITY: Lexington  
STATE: MA  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/471,869  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/08267  
FILING DATE: 22-JUL-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/095,734  
FILING DATE: 22-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/711,334  
FILING DATE: 06-JUN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/367,894  
FILING DATE: 19-JUN-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US90/03451  
FILING DATE: 18-JUN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US89/02962  
FILING DATE: 07-JUL-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/361,944  
FILING DATE: 05-JUN-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/223,089  
FILING DATE: 22-JUL-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/216,390  
FILING DATE: 07-JUL-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/163,546  
FILING DATE: 03-MAR-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US88/00614  
FILING DATE: 29-FEB-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/020,451  
FILING DATE: 02-MAR-1987  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: WH193-11MA2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-861-6240  
TELEFAX: 617-861-9540  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:



---

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2000, 02:46:36 ; Search time 1168.17 Seconds  
(without alignments)  
5292.534 Million cell updates/sec

Title: US-09-246-451-1  
Perfect score: 1402  
Sequence: 1 ctgcagatcgtatccgcgt.....gcgactaccaaagcgtata 1402

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues

Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*

44: gb\_est25:\*  
45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: em\_est20:\*  
50: em\_est21:\*  
51: em\_est22:\*  
52: gb\_est30:\*  
53: gb\_est31:\*  
54: gb\_est32:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: em\_est35:\*  
83: em\_est36:\*  
84: em\_est37:\*  
85: gb\_est48:\*  
86: gb\_est49:\*  
87: gb\_est50:\*  
88: gb\_est51:\*  
89: gb\_est52:\*  
90: gb\_est53:\*  
91: gb\_est54:\*  
92: gb\_est55:\*  
93: gb\_gss1:\*  
94: gb\_gss2:\*  
95: gb\_gss3:\*  
96: gb\_gss4:\*  
97: em\_gss1:\*  
98: em\_gss2:\*  
99: em\_gss3:\*  
100: em\_gss4:\*  
101: gb\_gss5:\*  
102: gb\_gss6:\*  
103: em\_gss5:\*  
104: em\_gss6:\*  
105: gb\_gss7:\*  
106: gb\_gss8:\*  
107: gb\_gss9:\*  
108: em\_gss7:\*  
109: em\_gss8:\*  
110: gb\_gss11:\*  
111: gb\_gss10:\*  
112: em\_gss9:\*  
113: em\_gss10:\*  
114: em\_gss11:\*  
115: em\_gss12:\*  
116: gb\_gss12:\*



KEYWORDS	EST
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eumariota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 408)
AUTHORS	Marras, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.
TITLE	The Mashu-HMI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	Contact: Marra M/Mouse EST Project Mashu-HMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@wustl.edu This clone is available royalty-free through LINDL; contact the IMAGE Consortium (info@image.lindl.gov) for further information. MG1:288584 Seq primer: -28M13 rev2 from Amersham High quality sequence stop: 373. Location/Qualifiers
FEATURES	1..408 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="IMAGE:477840" /clone_1id="Soares mouse embryo NDME13.5 14.5" /sex="unknown" /tissue_type="embryo" /dev_stage="13.5-14.5dpc total fetus" /lab_host="DH10B" /note="Vector: p7T73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer 15', TGTATCCATCTGTAAGTGGGCGGCGGCGGAAATTTTTTTTTTTTTTTTTTTT T 373), on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2 1]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Boudado."
BASE COUNT	76 a 127 c 123 g 82 t
ORIGIN	
Query Match	3.0%; Score 42.6; DB 20; Length 408;
Best Local Similarity	48.5%; Pred. No. 0.51;
Matches 117; Conservative 0; Mismatches 124; Indels 0; Gaps 0;	
OY	456 cccaccatgatgatgacccgcccagagcgccagcttgcgtgcgtggccaaccagtggt 515
Db	333 CCCACCTCTCTGGGGCGAGGCAATTTAGTGCACGCCGGGGTGCTGTGGCTGTGAAGGCAC 274
OY	516 ggcacatgcggttgatgaataagtcgtgagaacccggaatccagagctgcgtcgcgtgac 575
Db	273 TACCTCGAGGTGAGAGTGTCCAGGCGGACGAGCTGTCCAGGGAAGAGGCGGTGAGCTGCAC 214
OY	576 gagagctgcgcgcgcgaagagacagtgcacactaccagagagactacgcgcgaacctcccg 635
Db	213 TACCACTGCTCTATCCAGACAGAGTGCGCCAGCCCTCTGTGGAAGAGGAATCCACTATA 154
OY	636 ataccacatctatcgtcgtcgcaggtcaccgcgaagaagaatataccgcgacttgaaatac 695
Db	153 CTGGGACACAGGCTTCTCTGCTGACCTTGGCAAGCAGGTGTTTAAAGCAGGTACCAAGAAATGC 94
OY	696 c 696

	DB	93 C	93	
	RESULT	3		
	LOCUS	W45899/c		
	DEFINITION	W45899 411 bp mRNA EST 23-MAY-1996 mc78d12.r1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA clone IMAGE:354647 5', mRNA sequence.		
	ACCESSION	W45899		
	VERSION	W45899.1 GI:1330606		
	KEYWORDS	EST.		
	SOURCE	house mouse. Mus musculus		
	ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
	REFERENCE	1 (bases 1 to 411)		
	AUTHORS	Marras,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubouque,T., Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Thasing,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.		
	TITLE	The WashU-HMI Mouse EST Project		
	JOURNAL	Unpublished (1996)		
	COMMENT	On Apr 14, 1993 this sequence version replaced gi:837668. Contact: Marra M/Mouse EST Project WashU-HMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMGE Consortium (info@image.llnl.gov) for further information. MGI:226447 Seq primer: ETPrimer. Location/Qualifiers 1..411 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone_image="354647" /clone_idb="Soares mouse embryo NBME13.5 14.5" /sex="unknown" /tissue_type="embryo" /dev_stage="13.5-14.5dp total fetus" /lab_host="DH10B" /note="Vector: pRT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer 15' TGTTACCAATCTGAAGTGCGGAGCCGCCGGAATTCTTTTTTTTTTTTTTTTTTT T 3'), on equal amounts of mRNA from 2 13.5dp and 2 14.5dp embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2 ]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonafide."		
	BASE COUNT	78 a 127 c 124 g 82 t		
	ORIGIN			
	Query Match	3.0%;	Score 42.6;	DB 91; Length 411;
	Best Local Similarity	48.5%;	Pred. No. 0.51;	
	Matches 117; Conservative	0;	Mismatches 124;	Indels 0;
	Gaps	0;		
	456 ccaccactgatgataccggccagacgcaggcttcgtgctgagcacaagaagtgtt	515		
	Db 332 CCCACTCTCTGGGGCCAGGCATTTTCAGTCCACGGGGGTGCTGGGCTGTAAAGCAC	273		
	516 ggcatgccgttgtagataagctggaagacagatcacagaagctgctctgcctgtaac	575		

Db	272	TACCGCAGGTGAGAGTCTCCAGGSGCGAGCAGTGCAGGAGGAAGGCGCGTAGTGCAC	213
QY	576	gagagccctggccgcgcgaagaaatgcaacttcaccggagataagccgaaccctcccg	635
Db	212	TACCAACGCTGCTATCATCAAGCAGAGTGCGCCAGCCCTCTGTGGAAGAAGCAATCCACTATA	153
QY	636	atagcacttcattcatgtctgtcgcaggtctaccggaagaagatatacccgacttgaatatc	695
Db	152	CTGGGACACAGGCGTTTCTCTGCTGACCTTGCGCAAGCAGGTGTTTAGCAGAGTACCAGAGAGTC	93
QY	696	c 696	
Db	92	c 92	
RESULT	4		
CNS017SY			
LOCUS			
DEFINITION	CNS017SY 1101 bp DNA GSS 26-JUL-1999		
ACCESSION	Drosophila melanogaster genome survey sequence SP6 end of BAC		
VERSION	BACN37L08 of DrosBAC library from Drosophila melanogaster (fruit		
KEYWORDS	fly), genomic survey sequence.		
SOURCE	ALI08460		
ORGANISM	ALI08460.1 GI:5628764		
	GSS.		
	fruit fly.		
	Drosophila melanogaster		
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;		
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
REFERENCE	1 (bases 1 to 1101)		
AUTHORS	Genoscope.		
TITLE	Direct Submission		
JOURNAL	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :		
	BP 191 91006 EVRY cedex - FRANCE (E-mail : segre@genoscope.cns.fr		
	- Web : www.genoscope.cns.fr)		
	Determination of this BAC-end sequence was carried out as part of a		
	collaboration with the European Drosophila Genome Project (EDGP) -		
	http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC		
	library (Dros BAC) was made by Alain Billaud at CEPH (Centre		
	d'Etude du Polymorphisme Humain) with funding provided by a MRC		
	project grant. The DNA was prepared from embryos by Alain Bucheton		
	and Genevieve Payan. It has been constructed in the vector		
	pbeloBAC11.		
FEATURES			
SOURCE	location/qualifiers		
	1..1101		
	/organism="Drosophila melanogaster"		
	/plasmid="pbeloBAC11"		
	/db_xref="taxon:7227"		
	/clone_lib="DrosBAC"		
	/clone="BACN37L08"		
	/note="end : SP6"		
BASE COUNT	254 a 176 c 160 g 152 t 359 others		
ORIGIN			

	Query Match	Best Local Similarity	2.9% 13.2%	Score 41	DB 123	Length 1101
	Matches 51	Conservative 175	Mismatches 161	Indels 0	Gaps 0	
OY	256	atctgtctgcgcgcgtcgaagagccttgcgcagttcttcaagaatcaacgtaaccggatc	315			
Dh	709	AKMCCSSSSSSSSSCSAStressSsRKGMVSSCACSSGSGAASACGASGAGGSGRGK	768			
OY	316	tgtgtgtgactcgtcgcgaacgcgcgacactgtatgcgcactcgcgcgaactgactcgt	375			
Dh	769	GSGCCASGVCGAACSSSASCSSAMCGVSSGSSCASGCGGVSSCSAVSASSVMS	828			
OY	376	agcgctatgaagatattaccgcacatttcacaggaatgcgcgtatccctctgttgaaacgc	435			
Dh	829	KVASVVASCAVSSMSGMVSSSCRSSVASSVSAASVSSSVSSSVSAASVMS	888			
OY	436	gcgaagcctacattcattccacactcgtatgatacgccgcgaagcgcgcagttcgt	495			

Db 889 SSSASASMAVAAAAA/VASVSASVSSSSCCSSSSASVVSASVAAASASVSSSS 948

Qy 496 cgcgtgcaccaaccaagtggtgtgcataccgcggtgtgtgataagctgtgaagacccgataccagg 555

Db 949 SSVSTSSASVSVSASAVSMASVSSSSASSSSVSVSVVAASAAAAAASSSSSSS 1008

Qy 556 agctgcgcctctgcctgcatactgaagcctgcgcgccgaagaagacagtgcaattccaccagg 615

Db 1009 ASAAVAVVSASSSSSASSSSSSSASSSSVSSSCSVSVSSSVSSSVVAVASASASASV 1066

Qy 616 actacgcgaacccctccgatacca 642

Db 1069 SSVAAAVAAAAAASASASASVAA 1095

RESULT	5
LOCUS	AI388899
DEFINITION	AI388899 547 bp mRNA EST 28-JAN-1999
ACCESSION	GH19957.5prime GH Drosophila melanogaster head pot2 Drosophila
VERSION	AI388899
KEYWORDS	melanogaster cDNA clone GH19957 5prime, mRNA sequence.
SOURCE	AI388899.1 GI:4202919
ORGANISM	EST.
	fruit fly.
	Drosophila melanogaster
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 547)
AUTHORS	Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C.,
	Brokstein, P., Lewis, S. and Rubin, G.M.
TITLE	BDGP/HMI Drosophila EST Project
JOURNAL	Unpublished (1997)
COMMENT	Contact: Harvey, D.
	G. M. Rubin-Molecular and Cell Biology
	University of California Berkeley
	539 LSA, Berkeley, CA 94720-3200, USA
	Fax: 510 643 9947
	Email: <a href="http://www.fruitfly.org/EST">http://www.fruitfly.org/EST</a> , <a href="mailto:est@fruitfly.berkeley.edu">est@fruitfly.berkeley.edu</a>
	Plate: 199 row: E column: 9
	High quality sequence stop: 540.
FEATURES	Location/Qualifiers
source	1..547
	/organism="Drosophila melanogaster"
	/db_xref="taxon:7227"
	/clone="GH19957"
	/clone_lib="GH Drosophila melanogaster head pot2"
	/sex="male and female"
	/dev_stage="adult"
	/lab_host="DH5 - alpha"
	/note="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2:
	XhoI; Sized fractionated cDNAs were directly ligated into
	pOT2. Plasmid cDNA library."
BASE COUNT	135 a 153 c 171 g 88 t
ORIGIN	

Query Match	2.9%;	Score 40.6;	DB 39;	Length 547;
Best Local Similarity	49.3%;	Pred. No. 1.8;		
Matches 106;	Conservative 0;	Mismatches 109;	Indels 0;	Gaps 0;
QY 523	cagtgatggaataagctggaagaaccggaatccagaagctlgacctgctcgctgatacgaagcc	582		
Db 260	CGGAGCTGAACCAAGGTGTGGCCACCGCATCGAGATCATGCGCGGCGCATGACGCCAGAGCC	319		
QY 583	tggcgccgcaagaagcagtlgcaacttaccggaagactacgcgcgaacccttcccgatacga	642		
Db 320	GGCTGATGCTCGGGAGATATCCAGTCCGCTTCGATGGCTACACAAACAGCACTCTCATCA	379		
QY 643	tcttcattgctgctgcagagcttaccggaagaagatatccgcaacttgaaataacctaaagc	702		
Db 380	CCATGATGAGGTCTGGCGCACGCCGCTGGGAGGAGACCCAGAAACACTGTATGCGCATGAGAGC	439		



Oy	664	taccggaag	672
Dd	218	TGACTGCAG	226

  

RESULT	8		
LOCUS	W71279	461 bp	mRNA
DEFINITION	me1c10.r1 Soares mouse embryo NM0E13.5 14.5 Mus musculus cDNA		
ACCESSION	W71279		
VERSION	W71279.1	GI:1380017	
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Euairytia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 461)		
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisell,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theisinger,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.		
TITLE	The WashU-HMI Mouse EST Project		
JOURNAL	Unpublished (1996)		
COMMENT	On Jan 5, 1998 this sequence version replaced gi:1268847. Contact: Marra M/Mouse EST Project WashU-HMI Mouse EST Project Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouse@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium ( <a href="mailto:info@image.llnl.gov">info@image.llnl.gov</a> ) for further information. MGI:240938 Seq primer: ETPrimer High quality sequence stop: 343.		

  

FEATURES	
SOURCE	

```

1..461
    location/qualifiers
      /organism="Mus musculus"
      /strain="C57BL/6J"
      /db_xref="taxon:10090"
      /clone_image="IMAGE:389106"
      /clone_lib="Soares mouse embryo NM0E13.5 14.5"
      /sex="unknown"
      /issue_type="embryo"
      /dev_stage="13.5-14.5dp total fetus"
      /lab_host="DH10B"
      /note="Vector: pRT73D-Pac (Pharmacia) with a modified polynuker; Site.1: Not I - oligo(dT) primer [5', TGTTTCACAATCTGAAGTGCGAGCCGGAGAAATTGTGTGTGTGTGTGTGTGTGT T 3'], on equal amounts of mRNA from 2 13.5dp and 2 14.5dp embryos (total RNA provided by Minoru Ko, Wayne State Univ., from 2 ] double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT          91 a   148 c   147 g           75 t
ORIGIN

```

  

Oy	468	gatcgccgcggcaggccagcttcgtgcgcgtgcgccaccacaagtgttgcatgcgcggtg	527
Dd	129	GAGCGCGCTTGTGTGAAGAAGCATGTTAATCGCGGGCACCTCACTCTTGTTCGGGGGACACAC	188

  

Query Match	2.9%	Score 40.2;	DB 91;	Length 461;
Best Local Similarity	49.8%;	Pred. No. 2.3;		
Matches 102;	Conservative 0;	Mismatches 103;	Indels 0;	Gaps 0;

Oy	528	gtgataagctcggaggaacccgatccagaagctgcgtctctgcatatgatagacctgcgc	587
Db	189	GCGGGGCTGTGGAAACCGAAGCAGAGAGAAAAGGCCTCGATGTCGTCGGCGG	248
Oy	588	cgcgaagacagctgcaactcacccaggagactacgccaccctcccgatacgaatttc	647
Db	249	CCAGAAGAACTGTGGGAAGCGCGTGTGGGGGCCGCCAGCGCTTTCAAGCCCGCACTGGC	308
Oy	648	atcgtgctcgcaagtctacgggaag	672
Db	309	ACAAGACTGTGATTACTGACTGCAG	333
RESULT	9		
LOCUS	AA536622	531 bp mRNA EST	29-JUL-1997
DEFINITION	vJ71e04.f1 Knowles Solter mouse blastocyst B1 Mus musculus CDNA clone IMAGE:934494 5', mRNA_sequence.		
ACCESSION	AA536622		
VERSION	AA536622.1 GI:2282615		
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Materston,R.		
TITLE	The WashU-HHMI Mouse EST Project		
JOURNAL	Unpublished (1996)		
COMMENT	Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of MedicineP 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:539414		
FEATURES	High quality sequence stop: 168.		
SOURCE	Location/Qualifiers		
	1..531		
	/organism="Mus musculus"		
	/strain="B6D2 F1/J"		
	/db_xref="taxon:10090"		
	/clone="IMAGE:934494"		
	/clone_lib="Knowles Solter mouse blastocyst B1"		
	/tissue_type="blastocyst"		
	/dev_stage="embryo (pre-implantation)"		
	/lab_host="DH10B"		
	/note="Organ: embryo; Vector: pSPORT; Site_1: NotI; Site_2: SalI; Cloned unidirectionally from mRNA prepared from 800 blastocysts Primer: SalIdTn) 5'-CGGTGACCGTCGACCGCTTTTCTTTTTTTT-3'. cDNAs were cloned into the NotI/SalI sites of a pSPORT vector (Life Technologies). Two different size selections: B1 (larger inserts) and B3."		
BASE COUNT	99 a 139 c 217 g 76 t		
ORIGIN			
Query Match	2.9%; Score 40.2; DB 27; Length 531;		
Best Local Similarity	49.8%; Pred. No. 2.3;		
Matches 102; Conservative	0; Mismatches 103; Indels 0; Gaps 0;		
Oy	468	gacgcgccgagcagcgcagatttcgtgcgctgcgcaccaacaagtgttgcatgcgcgltg	527
Db	13	GAGCGGCTTGTTGGTAAGCACTTAATCATCGGGGCGCACTACATCTCTTTGCGGGGCGACGAC	72

[illegible]

FEATURES	COMMENT
	<p>Determination of this BAC-end sequence was carried out as part of collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> The BDGP Drosophila melanogaster BAC library was prepared by Kazuhiro Osoegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <a href="http://bacpac.med.buffalo.edu/drosophila_bac.htm">http://bacpac.med.buffalo.edu/drosophila_bac.htm</a>.</p>
Location/Qualifiers	

```

source
1. .844
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_1b="RPCI.98"
/clone="BACR11p16"
/note="end : TET3"
261 a 112 c 92 g 35 t 344 others
BASE COUNT
ORIGIN

```

Query Match	2.9%	Score 40.2;	DB 122;	Length 844;
Best Local Similarity	14.9%	Pred. No. 2.6;		
Matches 57;	Conservative 157;	Mismatches 168;	Indels 0;	Gaps 0.

[illegible]

Db 624 YYSBSCSGCGSSSSSSSCSSSGSSSBYSOCBCBCCBGYCGCYBWCSSBSSTBBSVSYGSBS 565BS

Qy 323 gactgcgtcaacagcgagcaacttgatcgcgcacccgcgcgcgaactatgctcgtaaggcta 3622

Db 564 TTSGTGCKCBSSSSSBBCSSBSGCCCTCCTCSYCTCYSCCHYCCBSGCHBSYSCSGB 505S

Qy 383 tgaagattaccgcgaactttccacagcagtgccggttcatccctcgtgaagccgcgaagc 442

Db 504 TGBSSBSGCKGKGSKSKGGBKGGKGYCYCPTTCCBSBSBCCSSSSSTSKCCSGSS 445S

Qy 443 ctacagacttcatcccaactcgaatgcgacgcgcgcgaagcagcagcttgcgtgcgcgc 5022

Db 444 STKCGSPBCCGSSSKMTGSSSCBBSGSSSTGSSSKKYSYTSSSSSGCGTGNNTTGC 385S

Qy 503 caaccagatggtgcatgcgc 524

Db 384 VGSCTTTTKSSCTBBSGTSBS 363

RESULT	11
AA690552	
LOCUS	474 bp mRNA EST 16-DEC-1997
DEFINITION	vz24d07.r1 Barted mouse myotubes MPR48 muscle cDNA clone IMAGE:1164013.5; similar to U000815.000815
ACCESSION	MTO-INOSITOL-1-PHOSPHATE SYNTHASE.; mRNA sequence.
VERSION	AA690552
KEYWORDS	GI:2691488
SOURCE	EST.
ORGANISM	mouse mouse. Mus musculus

REFERENCE 1 (bases 1 to 474)  
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Gessel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellensberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Willson,R. and  
Waterston,R.  
TITLE The Mashu-HMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On Jan 19, 1998 this sequence version replaced gi:2286451.

wasnu-HMMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LNL ; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 MGI:629925  
 Seq primer: -28ml3 rev2 ET from Amersham  
 High quality sequence stop: 472.  
 Location/Qualifiers

```

source
1. .474
/organism="Mus musculus"
/strain="C3H"
/db.xref="taxon:10090"
/clone="IMAGE:1164013"
/clone.lib="Barstead mouse myotubes MPLR5"
/cell_line="C2C12"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTCACGAATCTCGAAGTGGAGCGCGCCCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[AAATGGGATCCCTGG], digested with Not I and cloned into the
Not I and Eco RI sites of the modified pT73D vector.
Library constructed by Bob Barstead. The C2C12 cell lines
(available from ATCC, catalog # CRL-1772) differentiates
rapidly, forming contractile myotubes and producing

```





[illegible]

OY	438	gaagctacgacttattccatcccagtaggttcgcgcccgagcagcgacttgctg	497
Db	644	SSTTTCCTCKKCKCVCYTCTYTTTTSBCCSCGGBTTBSCSSSSSGSTTTSSSSS	585
OY	498	ctggccaaccaagtgttgcatcgccg	524
Db	584	SSSKCGSTSSSSSTSGSGTSSCS	558
RESULT	15		
AV389986			
LOCUS	AV389986	411 bp	mRNA EST 21-JAN-2000
DEFINITION	AV389986 Chlamydomonas reinhardtii C9 Chlamydomonas reinhardtii		
ACCESSION	CDNA clone CM054f01_r,		mRNA sequence.
VERSION	AV389986		
KEYWORDS	AV389986.1 GI:6544202		
SOURCE	EST.		
ORGANISM	Chlamydomonas reinhardtii.		
	Chlamydomonas reinhardtii		
	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;		
	Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.		
REFERENCE	1 (bases 1 to 411)		
AUTHORS	Asmizli,E., Nakamura,Y., Sato,S., Fukuzawa,H. and Tabata,S.		
TITLE	A Large Scale Structural Analysis of cDNAs in a Unicellular Green Alga, Chlamydomonas reinhardtii. I. Generation of 3433 Non-redundant Expressed Sequence Tags		
JOURNAL	DNA Res. 6, 369-373 (1999)		
COMMENT	On Jul 8, 1999 this sequence version replaced gi:5422438. Contact: Yasukazu Nakamura The First Laboratory for Plant Gene Research Kazusa DNA Research Institute Yana 1532-3, Kisarazu, Chiba 292-0812, Japan Email: ynakamu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/. Location/Qualifiers		
FEATURES	source		
	1..411		
	/organism="Chlamydomonas reinhardtii"		
	/strain="C9"		
	/db_xref="taxon:3055"		
	/clone="CM054f01_r"		
	/clone_lib="Chlamydomonas reinhardtii C9"		
	/dev_stage="photoautotrophic growth"		
	/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"		
BASE COUNT	69 a 153 c 128 g 61 t		
ORIGIN			
Query Match	2.6%; Score 36.8; DB 62; Length 411;		
Best Local Similarity	48.6%; Pred. NO. 18;		
Matches 101; Conservative 0; Mismatches 107; Indels 0; Gaps 0;			
OY	1046	cgatggcgcgatctcaacctccgaltacgagttcatgvcgtgcaactgaagaagtga	1105
Db	197	CCGGGCCCTGCAGCCCCCTGTCCGAGAAGAGCTTACGCCAAGCTTGGAAMAAGAGACT	256
OY	1106	ccaagtcctgcgaaccggagatctgtctgscctgatgtagcgcgaaacgcctgcccgt	1165
Db	257	GAAAGACCCTGGAGAAAGCCCTTAAGACATGACAGAGGGGACACAGCCCGCCGCTGACCT	316
OY	1166	gaagctcgacttcagtcgcgcaaaagttlcaacaacaacaccttggcgacgacgatc	1225
Db	317	GAAAGGCACCAATGGAAGCGCACCAAGGCCCGCTTCGCAACTACGCTAAGGCGGCTGCT	376
OY	1226	gtgccttgccagcagcctggccccgcg	1253
Db	377	GTGCGGCAACGACGCTGCTGCCACCTG	404

Search completed: October 4, 2000, 05:46:18  
Job time: 10782 sec

Search completed: October 4, 2000, 05:46:18  
Job time: 10782 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 05:49:24 ; Search time 76.18 Seconds

(without alignments)  
128.722 Million cell updates/sec

Title: US-09-246-451-2

Perfect score: 2180  
Sequence: 1 TTETIOSNANLAPLPHPVPE.....IVSGVOALPLVWDPATTKAV 414

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : A\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2137	98.0	405	Y04128	Pseudomonas putida
2	1185	54.4	446	Y04126	Bacterial and mamm
3	365	16.7	396	R60777	Mycobacterium IV hyd
4	332	15.2	408	R77867	S. clavuligerus OR
5	330.5	15.2	587	W33274	S. fradiae tylosin
6	330	15.1	406	R11349	Cytochrome enzyme
7	329	15.1	411	W54389	Actinomadura hibis
8	324.5	14.9	398	R11585	Streptomyces prist
9	322	14.8	403	R11350	Cytochrome enzyme
10	316	14.5	412	R38309	Sequence of the p4
11	313.5	14.4	410	R51368	Protein containing
12	297.5	13.6	404	R14724	6-hydroxylase enco
13	190	8.7	398	R47521	Vitamin D hydroxyl
14	161	7.4	422	W36128	Daunomycin C-14 hy
15	161	7.4	474	W36132	N-terminal modifie
16	156	7.2	422	W00729	Daunorubicin 14-hy
17	145	6.7	518	W67616	A. nidulans phenyl
18	144	6.6	494	R62825	Human steroid-21-h
19	144	6.6	533	R15057	Cytochrome P450C25
20	129.5	5.9	1169	R76544	Mitochondrial cyto
21	129	5.9	512	W93216	Human cytochrome p
22	128	5.9	516	W67617	P. chrysogenum phe
23	127	5.8	508	W35711	Chrysanthum flavon
24	127	5.8	512	R72365	Human auxillary cy
25	127	5.8	512	R93172	Human cytochrome p
26	126	5.8	503	R72363	Human cytochrome p
27	126	5.8	503	R81464	Human derived cyto
28	126	5.8	503	R93170	Human cytochrome p
29	126	5.8	503	Y05202	Human CYP3A4 prote
30	126	5.8	512	R72366	Human auxillary cy
31	126	5.8	512	W00652	Cytochrome P4501A1
32	126	5.8	512	R93173	Human cytochrome p
33	124.5	5.7	493	R72362	Human cytochrome p

34	124.5	5.7	493	1	R81467	Human derived cyto
35	124.5	5.7	493	1	R93169	Human cytochrome p
36	124	5.7	524	1	P70577	Rat hepato-cytochr
37	121.5	5.6	523	1	R59291	Rat liver cytochro
38	121.5	5.6	898	1	P61030	Entire coded seque
39	121.5	5.6	898	1	P61082	Complete translati
40	121.5	5.6	899	1	P61056	Translation of pla
41	121.5	5.6	1144	1	P81334	Expression prod. o
42	121.5	5.6	1150	1	P81335	Expression prod. o
43	121.5	5.6	1150	1	P81337	Expression prod. o
44	121.5	5.6	1162	1	P81336	Expression prod. o
45	121	5.6	512	1	R72364	Human auxillary cy

ALIGNMENTS

RESULT ID	1	Y04128	Y04128 standard; Protein; 405 AA.
AC	Y04128:		
DT	11-JUN-1999 (first entry)		
DE	Pseudomonas putida cytochrome P450 protein P450-cam.		
KW	Bacterial; mammalian; cytochrome P450; chimeric; fusion protein;		
KM	oxidise; hydrocarbon; carbon-hydrogen bond; hydroxylating;		
OS	bioremediation; environmental pollutant.		
PN	Pseudomonas putida.		
PD	MO9908812-AL.		
PE	25-FEB-1999.		
PF	17-AUG-1998; U16979.		
PR	20-AUG-1997; US-056754.		
PA	(UYRP ) UNIV ROCHESTER.		
PI	Jones JP, Shimoi M;		
DR	WPI, 99-190131/16.		
DR	N-PDB; X19926.		
PT	New P450 fusion proteins - comprising a portion of a bacterial		
PT	cytochrome P450 protein and a portion of a mammalian cytochrome P450		
PS	protein		
CC	Disclosure: Page 12-13: 51pp: English.		
CC	The present invention describes a fusion proteins comprising a portion		
CC	of a bacterial cytochrome P450 protein and also a portion of a mammalian		
CC	cytochrome P450 protein. The fusion protein can oxidise hydrocarbons or		
CC	any compound having a carbon-hydrogen bond. The fusion protein can be		
CC	used for hydroxylating a compound to be oxidised. It can also be used in		
CC	the bioremediation of an environmental pollutant. Since the fusion		
CC	protein is soluble, it can be subject to structural elucidation by X-ray		
CC	crystallography for designing functional proteins. It can be readily		
CC	expressed in soil bacteria to facilitate bioremediation. The present		
CC	sequence represents Pseudomonas putida cytochrome P450 protein P450-cam		
CC	from the present invention.		
SQ	Sequence 405 AA:		
Query Match 98.0%; Score 2137; DB 1; Length 405;			
Best Local Similarity 100.0%; Pred. No. 9.2e-207; Indels 0; Gaps 0;			
Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
OY	10	NLAPLPHPVDEHLVFDEDMYNPSNLGAGVDEANAVLQESNVPLDWTNRKNGHMIATRGQ	69
OY	1	NLAPLPHPVDEHLVFDEDMYNPSNLGAGVDEANAVLQESNVPLDWTNRKNGHMIATRGQ	60
OY	70	LIREAEVDYHPSSECEFTIREAGEADFTPTSGMDPRPQGFALANOVVGMVYDKLEN	129
OY	61	LIREAEVDYHPSSECEFTIREAGEADFTPTSGMDPRPQGFALANOVVGMVYDKLEN	120
OY	130	RIEELACSLIESLRPOGQCFTEDEYAEPPRIIRIFMLLAGIPEEDIPHLKYLTPQMTPPDG	189
OY	121	RIEELACSLIESLRPOGQCFTEDEYAEPPRIIRIFMLLAGIPEEDIPHLKYLTPQMTPPDG	180
OY	190	SMFPAEKALDYLLPIIEORRQKPGTDAISIVANGVGRPTSDSEAKRMCGLLVVG	249
OY	181	SMFPAEKALDYLLPIIEORRQKPGTDAISIVANGVGRPTSDSEAKRMCGLLVVG	240

QY 250 LDTVNFSLFSMEFLAKSPENROELIERPERIPACEELLRRSLVADGRILTSDEYFHG 309  
 |||||  
 DB 241 LDTVNFSLFSMEFLAKSPENROELIERPERIPACEELLRRSLVADGRILTSDEYFHG 300  
 |||||  
 QY 310 VOLKKDQILLPOMLSGLDERENACPMHVDPSQKVSHTTFGSHLCLGHLARRRITIV 369  
 |||||  
 DB 301 VOLKKDQILLPOMLSGLDERENACPMHVDPSQKVSHTTFGSHLCLGHLARRRITIV 360  
 |||||  
 QY 370 TLKEMLTRIPDPSIAPCAQIOHKSGIVSGVQALPLVWDPAITKAV 414  
 |||||  
 DB 361 TLKEMLTRIPDPSIAPCAQIOHKSGIVSGVQALPLVWDPAITKAV 405  
 |||||

RESULT 2  
 Y04126  
 ID Y04126 standard; Protein: 446 AA.  
 AC Y04126:  
 DT 11-JUN-1999 (first entry)  
 DE Bacterial and mammalian chimeric cytochrome P450 protein.  
 KW Bacterial; mammalian; cytochrome P450; chimeric; fusion protein;  
 KM oxidase; hydrocarbon; carbon-hydrogen bond; hydroxylating;  
 KW bioremediation; environmental pollutant.  
 OS Synthetic.  
 PN MO9908012-A1.  
 PD 25-FEB-1999.  
 PF 17-AUG-1998; 016979.  
 PR 20-AUG-1997; US-056754.  
 PA (UYRP ) UNIV ROCHESTER.  
 PI Jones JP, Shimoji M;  
 DR WPI: 99-190131/16.  
 DR N-PSDB; X19916.  
 PT New P450 fusion proteins - comprising a portion of a bacterial  
 cytochrome P450 protein and a portion of a mammalian cytochrome P450  
 protein.  
 PS Claim 24: Page 6-8: 51pp; English.  
 CC The present sequence is a fusion proteins comprising a portion of a  
 bacterial cytochrome P450 protein and also a portion of a mammalian  
 cytochrome P450 protein. The fusion protein can oxidise hydrocarbons or  
 any compound having a carbon-hydrogen bond. The fusion protein can be  
 CC used for hydroxylating a compound to be oxidised. It can also be used in  
 the bioremediation of an environmental pollutant. Since the fusion  
 CC protein is soluble, it can be subject to structural elucidation by X-ray  
 CC crystallography for designing functional proteins. It can be readily  
 CC expressed in soil bacteria to facilitate bioremediation.  
 SQ Sequence 446 AA;

Query Match 54.4%; Score 1185; DB 1; Length 446;  
 Best Local Similarity 57.6%; Pred. No. 6, 6e-111;  
 Matches 253; Conservative 33; Mismatches 99; Indels 54; Gaps 9;

QY 10 NLAPLPVPEHVLVPFDYMNPSNLSAGVOEAMAVLOESNVPLVWTRCGHMTATRGQ 69  
 |||||  
 DB 1 NLAPLPVPEHVLVPFDYMNPSNLSAGVOEAMAVLOESNVPLVWTRCGHMTATRGQ 60  
 |||||  
 QY 70 LIREAVEDYRHFSSECPFIPREAGEAYDIFPTSMDDPEQROFRALANQVGMVVDKLEN 129  
 |||||  
 DB 61 LIREAVEDYRHFSSECPFIPREAGEAYDIFPTSMDDPEQROFRALANQVGMVVDKLEN 120  
 |||||  
 QY 130 RIDEACSLIESLRPOGOCFTEDYAEPPRIRIFMLAGLPREDIDHLKYLTDQMTRPDG 189  
 |||||  
 DB 121 RIDEACSLIESLRPOGOCFTEDYAEPPRIRIFMLAGLPREDIDHLKYLTDQMTRPDG 180  
 |||||  
 QY 190 SMFFAKKALDYLLPIITIOROKPGT-----DAISIYANGOVGRP--ITSDEAKRM 241  
 |||||  
 DB 181 SMFFAKKALDYLLPIITIOROKPGT-----DAISIYANGOVGRP--ITSDEAKRM 240  
 |||||  
 QY 242 CGLLVGLDVTYVNFSLFSMEFLAKSPENROELIERPERI-----PA 283  
 |||||  
 DB 241 AVDLFGAGTETSTLTLYALVALLLKKHREVTAKVQEEIERVIGNRSPQMDRSHMPTTDA 300  
 |||||  
 QY 284 ACEELLRRSLVADG--RLITSDEYFHGVOLKKGDOILL-----POMLSG- 326  
 |||||

DB 301 VHAENVORYTDLPTSLPHAVTCIDIKFRNRYLIPKFTTILISTSVLHDNKEFPNEMFDPH 360  
 |||||  
 QY 327 --LDERENACPMHVDPSROKVSHTTFGSHLCLGHLARRRITIVTLKEMLTRIPDPSIA 384  
 |||||  
 DB 361 HFLDEGNN-----FKSKY-FMPPSACKRICGEGALGEMELFLTLTSLONFNKLISLV 412  
 |||||  
 QY 385 PGAQIOHKSGIVSGVQALP 403  
 |||||  
 DB 413 DPKNLD-TTPVNVGFAFSVP 430  
 |||||

RESULT 3  
 R60777  
 ID R60777 standard; Protein: 396 AA.  
 AC R60777:  
 DT 21-JUN-1995 (first entry)  
 DE Mycinamicin IV hydroxylating protein.  
 KW Mycinamicin; hydroxylase; macrolide; antibiotic; Micromonospora;  
 OS Micromonospora griseorubida.  
 PN J06253853-A.  
 PD 13-SEP-1994.  
 PF 09-MAR-1993; 047638.  
 PR 09-MAR-1993; JP-047638.  
 PA (ASAH ) ASAH KASEI KOGYO KK.  
 DR WPI: 94-328997/41.  
 DR N-PSDB: Q73674.  
 PT DNA encoding a protein having mycinamicin IV hydroxylating  
 activity - for prodn. of mycinamicin, a macrolide antibiotic  
 PS Claim 1: Page 12-14; 23pp; Japanese.  
 CC The amino acid sequence of a protein having mycinamicin IV hydroxylase  
 CC (MH) activity. The gene encodes a protein of 396 a.a. The DNA was  
 CC obtained from the macrolide-producing bacterium Micromonospora  
 CC griseorubida A11725CN3/PT5507. The gene was isolated from the plasmid  
 CC pT5507. The protein encoded by this plasmid can be used to produce  
 CC mycinamicin IV in PT5507-deficient Micromonospora strains.  
 SQ Sequence 396 AA;

Query Match 16.7%; Score 365; DB 1; Length 396;  
 Best Local Similarity 28.5%; Pred. No. 1, 8e-28;  
 Matches 105; Conservative 57; Mismatches 168; Indels 38; Gaps 8;

QY 63 WINTRGQLIREAVEDYRHFSSECPF-----PREAGEAYDIFPTSMDDPEQROFRA 113  
 |||||  
 DB 43 WLVTYR-----YEDYRAVLGDSGRFVRCPSMTDRERTREPMYKGLSMDPREHRLRR 95  
 |||||  
 QY 114 LANQVGMVVDKLENRIODEACSLIESLRPOG-CNFTEDYAEPPRIRIFMLAGLPREE 172  
 |||||  
 DB 96 LVVKAFTTARAEISLRPRAREIAHELVDOMATGQPADLVAMFARQLPVRVYICELGVP 155  
 |||||  
 QY 173 DIPHLKYLTDQMTRPDGS-----TFAEKALDYLLPIITIOROKPRTDAISIV 223  
 |||||  
 DB 156 D-----HDFTRMSGALSTAEVTAEMOEAEQAVYMGDLIDRRKEPTDLDVSL 208  
 |||||  
 QY 224 ANQOVNGRPITTSDEAKRMGCLLVGLDVTYVNFSLFSMEFLAKSPENROELIERPERIPA 283  
 |||||  
 DB 209 VQARQDDSLSEBELLDIAIGLVAGYESTTQIOAFVLLMTRPRLRQGLDRPELIS 268  
 |||||  
 QY 284 ACEELLRRSL--VADGRILTSDEYFHGVOLKKGDOILLPOMLSGLDERENACPMHVD 340  
 |||||  
 DB 269 AVEELTRWVPLGVTGAFPRAYVEDVTLRGVTTIACGEPVLASTGANRDOAQFPDADRIV 328  
 |||||  
 QY 341 SRQVSHHTFGSHLCLGHLARRRITIVTLKEMLTRIPDPSIA--PGAQIOHKSG-IVSG 398  
 |||||  
 DB 329 DRTPNQHLGEGHVNHLGAPLAREVLALEVLQRLPGRIGRIGPDTLRMSEGMLRG 388  
 |||||  
 QY 399 VQALPLVM 406  
 |||||  
 DB 389 PLELPLVM 396  
 |||||

RESULT 4

R77867  
ID R77867 standard; Protein: 408 AA.  
AC R77867;  
DT 13-NOV-1995 (first entry)  
DE S. clavuligerus OREF10 product.  
KM Clavulanic acid; clavulanate; antibiotic; beta-lactamase-inhibitor.  
OS Streptomyces clavuligerus.  
PN CA2108113-A.  
PD 09-APR-1995.  
PF 08-OCT-1993; 108113.  
PR 08-OCT-1993; CA-108113.  
PI (UNAL-) UNIV ALBERTA.  
PI AIdoo KA, Jensen SE, Paradkar AS;  
DR WPI: 95-207301/28.  
DR N-PSDB: Q91580.  
PT Clavulanic acid biosynthesis enzymes and corresp. DNA - useful for  
PT biosynthesis of the antibiotic in Streptomyces hosts which do not  
PT naturally produce clavulanate  
PS Clam 32; Fig.19; 41pp; English.  
CC A 15 kb fragment S. clavuligerus NRRL 5741 genomic DNA (Q91580),  
CC extending downstream from pcbC, included 10 ORFs encoding the  
CC enzymes required for clavulanate biosynthesis. The OREF10  
CC product (R77867) showed high similarity to cytochrome P450-type  
CC enzymes from other Streptomyces spp.  
SQ Sequence 408 AA;

Query Match 15.2%; Score 332; DB 1; Length 408;  
Best Local Similarity 29.0%; Pred. No. 3.9e-25;  
Matches 122; Conservative 59; Mismatches 185; Indels 54; Gaps 18;

QY 16 PHEPHELVDFEDMYNNSAGVQAMAVLQESNVPDLVTRCNGH-WIATRGOLIREA 74  
DB 14 PAFPMRRCVPD---PPQLAGLRQKASRYT-----LM---DGSQVWLVTSHACARAY 62  
QY 75 YEDYRHS-SECFIP-----BEAGEYDFIPTSMDEPQROFRA-----LANQV 118  
DB 63 LGRFRFTAVTSAGFPMLTRTSQVLRANPESASFI--RMDPQHSRLRSLMTRDFLARRA 120  
QY 119 VGM-PPVVDKLENIQOLACSLIESLRPQOCNTEDEYAEFFPIRIMLAGLPREDIPHL 177  
DB 121 EALRPVAVREL---LDLILGLVGERP---VDLVAGLITTPVPSRYTLLFGADDREFFI 174  
QY 178 K---VLTDOMTPRPGSMTFAEAKELDYLIPIEORROKPGTDIAISVANGVNGRP 233  
DB 175 EDRSAVLIDRGTYPE---QVAKARDELQGLRLVEERIEENPCTDLISRVTIOVPRGHL 231  
QY 234 TSDAARMCGLLLVGGLDVTNVNLSFSMEFLAKSPENROELIERPERIPAAECLELRFES 293  
DB 232 RVEEMVPMCRLLLVAGHGTTTSQASLSLSTLTDPELAGRLTEDPALLPKAVBELLRFS 291  
QY 294 LVADG--RLITSYEFHGVQKKGDQILLPQMLSGIDERENACPMNVDSRQVSHRTTG 351  
DB 292 IYONGIARAAVEDVDLDVLRAGEGVVLSLGSNDETVLPDRDVRDARRHLAFG 351  
QY 352 HGSHLICIGHLAR---REIIVTKEMVLTIRIPDSIA--PGAIOHKSGIYS--GVQALPLW 406  
DB 352 HGMHQLGOMLAVLELEILAAVLRMM---PCARLAVPEELDFRHEVSSYGIGALPVTW 408

RESULT 5  
ID W33274  
W33274 standard; Protein: 587 AA.  
AC W33274;  
DT 17-FEB-1998 (first entry)  
DE S. fradiae tylosin biosynthesis gene product tyIH.  
KM Tylosin; biosynthetic gene product; production; antibiotic; tyIH.  
OS Streptomyces fradiae.  
PN US5672497-A.  
PD 30-SEP-1997.  
PF 21-DEC-1995; 575843.  
PR 12-MAY-1989; US-351350.  
PR 21-MAR-1986; US-842330.

PR 25-JUL-1986; US-890670.  
PR 24-FEB-1987; US-018237.  
PR 06-AUG-1991; US-742222.  
PR 28-JUL-1993; US-107232.  
PR 17-FEB-1994; US-198672.  
PR 21-DEC-1995; US-575843.  
PA (ELIL ) LILLY & CO ELI.  
PI Cox KL, Fishman SE, Hersberger CL, Seno ET;  
DR WPI: 97-488860/45.  
DR N-PSDB: T58686.  
PT DNA encoding Streptomyces fradiae tylosin biosynthesis gene products  
PT - for increasing tylosin production in Streptomyces spp.  
PS Clam 20; Columns 21-38; 38pp; English.  
CC The present sequence is the Streptomyces fradiae tylosin  
CC biosynthetic gene product tyIH, useful to increase the production  
CC of the antibiotic tylosin in Streptomyces spp. .  
SQ Sequence 587 AA;

Query Match 15.2%; Score 330.5; DB 1; Length 587;  
Best Local Similarity 29.3%; Pred. No. 9.7e-25;  
Matches 105; Conservative 62; Mismatches 170; Indels 21; Gaps 11;

QY 63 WIATRGOLIREAYEDYR--HFSSECFIPREAGEA-YDFIPTSMDEPQROFALANQV 118  
DB 143 WLISRODHNVALADPRVSIH-PAKLPRLSPSDGEASRSILTLDDPRHGALRGHPIPE 201  
QY 119 VGMPPVVDKLENIQOLACSLIESLRQG--OCNTEDEYAEFFPIRIMLAGLPREDIPHL 177  
DB 202 FGRLRVRELSPSEQIVTGLDLDLTARGDEADLADFALEPMATQVICRLDIYERBDYF 261  
QY 178 KYLTDOMTPRPGSMTFAEAKELDYLIPIEORROKPGTDIA--SVANGVNGRPITS 235  
DB 262 QERTEDQTRAAEEALELLELDYLDRLISKTGCHESGDMLGMYA--QARGGGLSH 319  
QY 236 DEAKRMCGLLLVGGLDVTNVNLSFSMEFLAKSPENROELIERPERIPAAECLELRFES 295  
DB 320 ADVLDNAVLLLAAGHETTASMTMSVULVLIQHPTAMRELIVNGLLPAGVADLLRYLS-I 378  
QY 296 ADG--RLITSYEFHGVQKKGDQILLPQMLSGIDERENACPMNVDSRQVSHRTTGHC 353  
DB 379 ADGLRSATVADIDIGITTTARGDLVFLAANRDEAVSEPEAFDIHRSARHNVAFGYG 438  
QY 354 SHLICIGHLARREIIVTKEMVLTIRIPDSIAP-----GAIOHKSGIYSGVQALPLW 406  
DB 439 PHOCLGONLARMELEVALCAVLERLP--ALRPITTVAGLKLKSDSA-VFQVYELPVAW 493

RESULT 6  
ID R11349  
R11349 standard; Protein: 406 AA.  
AC R11349;  
DT 05-JUN-1991 (first entry)  
DE Cytochrome enzyme P450SU1.  
KM Cytochrome P450; P450SU1; P450SU2; herbicide resistance.  
OS Streptomyces griseolus.  
PN M09103561-A.  
PD 21-MAR-1991.  
PF 27-AUG-1990; U04785.  
PR 11-SEP-1989; US-405605.  
PR 12-JAN-1990; US-464499.  
PR 23-AUG-1990; US-569781.  
PA (DUPQ ) DU PONT DE NEMOURS CO.  
PI Dean C, Harder PA, Leto KJ, Lichtner FT, Odell JT;  
PI O'Keefe DP, Omer CA, Komesser JA;  
DR WPI: 91-102077/14.  
DR N-PSDB: Q11126.  
PT DNA encoding cytochrome P450 enzymes - and electron donating  
PT iron-sulphur proteins, used to confer herbicide resistance to  
PT plants and microorganisms  
PS Clam 13; page 151; 224pp; English.  
CC This cytochrome P450 enzyme, P450SU1 is expressed alongside the iron  
CC sulphur protein Fes-B, by a DNA sequence contained in a recombinant

CC plasmid. Host Streptomyces species are transformed with the plasmid  
CC and are used to coat a plant seed to transform the plant. The res-  
CC ultant transformants are resistant to herbicides.  
CC See also 011127.  
SQ Sequence 406 AA:

Query Match 15.1%; Score 330; DB 1; Length 406;  
Best Local Similarity 26.3%; Pred. No. 6,2e-25;  
Matches 104; Conservative 74; Mismatches 193; Indels 24; Gaps 11;

QY 31 PSLMSGVQ--EMAVVLOESNVDLWTRKNGCH-WIATRGQILREAYEDYR-----79  
D 17 PNRSCPYDLPDGYAQLRDPGLHRYTLVDGQAMVWTKHEARKILGDPRLSSNRTD 76  
QY 80 HFSSECFPIR--REAGEAYDFIPTSMPPRQORFALANOVMPVVDKLENTIOELACS 137  
D 77 NFPAIYSPREAVAVESFOAF-----IGDDPRPHGTRRRKMTISEFTVKKIKGRPEVEEYVHG 132  
QY 138 LIESLRPGQ-QCNFTEDYAEPPFIRIFMLLAGLPEEDIPHLKYLTDQMRPPDSMTFAEA 196  
D 133 FLEMLAAGPTADLVGSOFALPVPSSVTCRLGLVYADHEFFQDASKRLVSTDAQSALTA 192  
QY 197 KEALYLYLPIIQOROKRGTDAI-SIVANGVNGRPITSDEAKRMCGLLVGGDLTVN 255  
D 193 RNDLAVYLDGLITQFTEPAGLVGALVADQLANGE-IDREELISTAMLLIGHETTAS 251  
QY 256 FLFSMEFLAKSPENHOELIERPERIPACEELLRRFSL--VADGRILTSYDFEHGVOLK 313  
D 252 MSLSYITLLDHEQYALRADRSLLVPGAVELLRYLALADINGAVATADIVEGHILIR 311  
QY 314 KGDQILLPQMLSGLDERENACPMHVDPSROKVSHTTEGHSHLCLGOHLARREIIVTLKE 373  
D 312 AGEGVIVVANSIARNDGTVEYEDPDALDIHRSARHHLAFEGVHQCLOGNLARLELYTLNA 371  
QY 374 WLTRIPDFSLA-PGAIOIHKSG-IVSGVALPLVW 406  
D 372 LMDRVPTLRLLAVPEQLVLRPGTTIGGVNELPYTW 406

RESULT 7  
W54389  
ID W54389 standard; Protein: 411 AA.  
AC W54389;  
DE 18-AUG-1998 (first entry)  
KW Actinomadura hibisca polyketide synthase protein 10.  
KW Multienzyme; infection; fungi; yeast; gram-positive bacteria; virus;  
KW dihydrobenzo(a)naphthacenequinone aglycon; antibiotic; pradimicin.  
OS Actinomadura hibisca.  
PN WO9611230-A1.  
PD 19-MAR-1998.  
PF 13-SEP-1996; U14791.  
PR 13-SEP-1996; WO-U14791.  
PA (BRIM) BRISTOL-MYERS SQUIBB CO.  
PI Dairi T, Oki T.  
DR WPI: 98-207391/18.  
DR N-PSDB: V26609.  
DR P-PSDB: W54389, W54390.  
PT Actinomadura polyketide synthase genes - useful for preparation of  
PT pradimicin  
PS Disclosure: Page 55-56; 71pp; English.  
CC The Actinomadura hibisca polyketide synthase proteins W54380-W54390 form  
CC a multienzyme complex. The enzyme is used for the biosynthesis of a  
CC dihydrobenzo(a)naphthacenequinone aglycon preferably a pradimicin which  
CC is an antibiotic useful against systemic fungal infections caused by  
CC Candida albicans, Aspergillus fumigatus and Cryptococcus neoformans. It  
CC is also active against a wide variety of fungi and yeasts, some  
CC Gram-positive bacteria and viruses.  
SQ Sequence 411 AA:

Query Match 15.1%; Score 329; DB 1; Length 411;  
Best Local Similarity 26.4%; Pred. No. 7.9e-25;

Matches 114; Conservative 61; Mismatches 199; Indels 58; Gaps 16;

QY 4 TIOSNMNLAPLRPHVEBHLVDFDWMYNP-----SNLSAGVQEMAVVLOESNVDLWTRCN 59  
D 9 TVDPREDVTTPAFPRPD-----DPFQPCENARKLRASDPKAVLP-----T 50  
QY 60 GGH-WIATRGQILREAYEDYRHFSSSEC---PFIPREAGEAYDFIPTSM-----MDPPE 107  
D 51 GDHAAVWTVRADVRFEVTS-D-RRFSKEAVTRPGAPR-----LIPMQRSKSLVMDPPE 102  
QY 108 QROFRALANOVMPVVDKLENTIOELACSLESIRPGQ-QCNFTEDYAEPPFIRIFMLLI 166  
D 103 HTRMRKIVSRKFTARRVEGKRAHVRDLTSGVDENVEHGPADLIHALLPLPVTVICEM 162  
QY 167 AGLPEEDIPHLKYLTDQMRPPDSMTFAEAK-----ALYDVLPIIQOROKRGTDAI 220  
D 163 LGVPEDRRFRFQWTRML-TIGAPALAOADEIKAAVGRIGRLAEILDKRTAARADDLL 221  
QY 221 SIVANGVNGRPITSDEAKRMCGLLVGGDLTVNPLFSMEFLAKSPENHOELIERPER 280  
D 222 SLISRAHADQ-GLSEBELLTFCGWTLLAGYHTTTAITSVYHLREPSRYARLRDPSG 280  
QY 281 IPACEELLRRSLVADG-----RILTSYDFEHGVOLKKGQDILLPQMLSGLDERENACPM 336  
D 281 IPAVAVELL-RVGOIGGAGAIRIAVEDVEVGGLTVRAGEAVILPNAANRPEVADPE 339  
QY 337 HVDPSROKVSHTTEGHSHLCLGOHLARREIIVTLKEMWLTRIPDFSLA-PGAIOIHKSGI 395  
D 340 ELDLGRITDNRHIALGCHITCLGAPLARLELOVYLTVERPALRIALIDDADITWRPGL 399  
QY 396 V-SGVQALPLVW 406  
D 400 AFAPRDALPLAW 411

RESULT 8  
W11585  
ID W11585 standard; Protein: 398 AA.  
AC W11585;  
DE 02-APR-1997 (first entry)  
KW Streptomyces pristinaespiralis Smbf gene product.  
KW Streptogramin B; antibiotic; biosynthesis; pristinaemycin;  
KW virginiamycin; pipicolinic acid; cyclodeamination; papa; snbf;  
KW pipA; 3-hydroxypipicolinic acid; hydroxylation.  
OS Streptomyces pristinaespiralis.  
PN WO9601901-A1.  
PD 25-JAN-1996.  
PF 04-JUL-1995; F00889.  
PR 08-JUL-1994; FR-008478.  
PA (RHON) RHONE-POULENC RORER SA.  
PI Barriere JC, Blanc V, Blanche F, Crouzet J;  
PI Debussche L, Paris JM, Thibaut D, Bamas-Jacques N;  
PI Dutruc-Rosset G, Famechon A;  
DR WPI: 96-097631/10.  
DR N-PSDB: T58555.  
PT New streptogramin B derivs. useful as antibiotics - produced by new  
PT mutants of Streptomyces having altered genes for streptogramin B  
PT biosynthesis  
PS Example 1; Page 113-114; 146pp; French.  
CC The papa gene of S.pristinaespiralis is involved in the biosynthesis  
CC of 4-dimethylamino-L-phenylalanine (DMPAPA), a precursor for  
CC streptidamycin IA. Upstream of the papa gene, on the complementary  
CC strand, is the snbA gene coding for 3-hydroxypipicolinic acid-AMP ligase.  
CC The region between these two genes was sequenced and two open reading  
CC frames were identified. The first (pipA) decodes to an amino acid  
CC sequence with homology to ornithine cyclodeaminase from Agrobacterium  
CC tumefaciens. The pipA gene product is likely to catalyse the cyclo-  
CC deamination of lysine, leading to production of pipicolinic acid.  
CC Mutations in the pipA gene were shown to affect pipicolinic acid  
CC synthesis but not the synthesis of 3-hydroxypipicolinic acid. The second  
CC open reading frame (snbf) could be decoded to give a product with  
CC homology to hydroxylases of the cytochrome P450 type.  
CC Disruption of the pipA and snbf genes can be used to produce

CC strains of *S.pristinaespiralis* which are unable to produce the  
CC antibiotic pristinamycin I but which may be able to produce new,  
CC modified forms of it.  
SO Sequence 398 AA;

Query Match 14.9%; Score 324.5; DB 1; Length 398;  
Best Local Similarity 29.0%; Pred. No. 2.1e-24;  
Matches 106; Conservative 59; Mismatches 159; Indels 41; Gaps 12;

QY 74 AYEDYRA-----FSSECPFIPREAGYDFIPISMPDEOROFALANOVGM 121  
DB 36 AFHFHRAVDLVASDPGVYSSQLSRPSQALSEQILSYIDPMHRTLRIVSQAFTP 95  
QY 122 PVYDKLENIROELACSLIESLRPGQC-NFTEDYAEPPPIRIFMLLGLPEEDIPIHLKYL 180  
DB 96 RYVADLEPRVETELAGQLIDAV--DGFDFDLVADRAYPLPVIAELLGVPRADTLFRSW 153  
QY 181 TDQMT-----PDGSMTEFAKEALYDYLPIIEORROKPGTDAISIVA 224  
DB 154 SDRMLQOVADPADMQGDADDEYQRLVKEPMAHAYLHDYDRARRANDLISALV 213  
QY 225 NGOVNGPPTISDEAKKRMCGLLVGLDTVVNFLSFSMEFLAKSPEHROELTERPER--IP 282  
DB 214 AARVEGERLTDEQIVFERGALLMAGHVSTSMLEGNFVLCIKDHP--RAEAAARADRSLLIP 271  
QY 283 AACCELLR-RPSLVADGRITLSDYEFHGVLKKGDQILLPQMLS-GUDERENACPMHVD 340  
DB 272 ALIEEYLRPRPIVMAVTTKDTVLAGTTIPAG-RNAVPSLSANHDEQVFTDPDLHL 330  
QY 341 SRQKVSHTTGGSHLCLGQHLARREIIVTLKEWLTIRIPDSIAPGAQIQ-HKSGIVSGV 399  
DB 331 ARBG-RQIARGHGHIYGLGAPLARLEGRALDELFDRFDPSPIDGAKLRHRCGLF-GV 388  
QY 400 QALPL 404  
DB 389 KNEPL 393

RESULT 9  
ID R11350 standard; Protein; 403 AA.  
AC R11350;  
DT 05-JUN-1991 (first entry)  
DE Cytochrome enzyme P450SU1;  
KM Cytochrome P450; P450SU1; P450SU2; herbicide resistance.  
OS Streptomyces griseolus.  
PN W09103561-A.  
PD 21-MAR-1991.  
PF 27-AUG-1990; U04785.  
PR 11-SEP-1989; US-405605.  
PR 12-JAN-1990; US-464499.  
PR 23-AUG-1990; US-569781.  
PA (DUPQ ) DU PONT DE NEMOURS CO.  
PI Dean C, Harder PA, Leto KJ, Lichtner FT, Odell JT;  
PI O'Keefe DP, Omer CA, Romesser JA;  
DR WPI: 91-102077/14.  
DR N-PSDB; Q11127.  
PT DNA encoding cytochrome P450 enzymes - and electron donating  
PT iron-sulfur proteins, used to confer herbicide resistance to  
PT plants and microorganisms  
PS claim 15; page 158; 234pp; English.  
CC This cytochrome P450 enzyme, P450SU2 is expressed alongside the iron  
CC sulphur protein res-A, by a DNA sequence contained in a recombinant  
CC plasmid. Host Streptomyces species are transformed with the plasmid  
CC and are used to coat a plant seed to transform the plant. The res-  
CC ultant transformants are resistant to herbicides.  
CC See also Q11126.  
SO Sequence 403 AA;

Query Match 14.8%; Score 322; DB 1; Length 403;  
Best Local Similarity 28.4%; Pred. No. 3.9e-24;

Matches 103; Conservative 64; Mismatches 166; Indels 30; Gaps 13;

QY 63 WIAHQGLIREAVDYDHFSECC-----PIPRRAGAYFIPIR--SMDPEQOFAL 115  
DB 52 WLTVRHODVRAVLGDDPR-FSADAHRTGPPPLTAGRGRIIGTNPFLRMDDPEHARLRML 110  
QY 116 NOVGMFVVDKLENIROELACSLIESLRP-OGQCNFTEDYAEPPPIRIFMLLGLPEEDI 174  
DB 111 TADPIYAKVEAMREYORLADDDLVDRMTGTSDADVTETAPLRPLSLVICLLGVPREDH 170  
QY 175 PHL---KYLTDMTRPDGSMTEFAKEALYDYLPIIEORROKPGTDAIS-IVANQVN 229  
DB 171 AFEQERSRVLTLTRSTPE---EVRAODELLEYLARLARTRKREPRDDAISRLVARGELD 227  
QY 230 GRPITSDEAKRMCGLLVGLDTVVNFLSFSMEFLASPEHROELTERPERIPACCELL 289  
DB 228 DTQIAT-----MGRLLVAVGGELEDDPQIATWGLLRNPOLARLRAEPALVKGAVBEL 282  
QY 290 RRFSLVADG--RILTSDEYFHVQLKKGQDILLPQMTSGDERENACP--MHVDFSRQV 345  
DB 283 RYLITVINGVPRIATDEVLLIGRTIAGEGVLC--MISSANROAEVPPGGDDLDVADAR 340  
QY 346 SHTTFGHSHLCLGQHLARREIIVTLKEWLTIRIPDSIA-PGAQIQHKSGI-VSGVALP 403  
DB 341 RHVAFGSGVHQCLOGPLARVELQIAETLLRLPLRLAVPHEEIPPRGDMAIYGVSLP 400  
QY 404 LVW 406  
DB 401 IAW 403

RESULT 10  
ID R38309 standard; Protein; 412 AA.  
AC R38309;  
DT 04-DEC-1993 (first entry)  
DE Sequence of the P450-soy protein.  
KM P450soy; soyC gene; oxidation; haem protein.  
OS Streptomyces griseus ATCC 13273.  
PN W09312236-A.  
PD 24-JUN-1993.  
PF 16-DEC-1992; U10885.  
PR 16-DEC-1991; US-807001.  
PA (DUPQ ) DU PONT DE NEMOURS & CO E I.  
PI Omer CA, Sariaslani FS, Trower MK;  
DR WPI: 93-214178/26.  
DR N-PSDB; Q45569.  
PT Constitutive expression of P450 SOY (SOYC) and ferredoxin soy  
PT (soyD) in Streptomyces - used for oxid. of organic chemicals  
PS Example; Figure 2; 45pp; English.  
CC Cytochrome P450soy was purified from *S. griseus* ATCC 13273. Two  
CC similar forms of P450soy were isolated. P450soy-delta, is derived  
CC from P450soy by in vitro proteolysis during isolation. One of the  
CC tryptic peptide fragments of cytochrome P450soy and of of the  
CC P450soy-delta protein were subjected to automated degradation. The  
CC NH2 terminal sequences are given in R38306 and R38307. A mixture of  
CC oligos that consist of possible DNA sequences that could encode the  
CC Aas FevHQC of the tryptic peptide was made. It consists of the  
CC sequences in Q43290-23. The oligo mixture was end-labeled and used  
CC to probe the EMBL4 library of *S. griseus* DNA. Hybridising plaques  
CC were isolated and a 4.8kb SacI DNA fragment was isolated from one  
CC clone that hybridised to the oligo probe mixture. As segment of the  
CC 4.8kb fragment was sequenced and found to contain an ORF. Within  
CC this ORF was a section that corresp. exactly to the AA sequence  
CC determined from the cytochrome P450soy tryptic peptide (see Q45569,  
CC R38309). The gene encoding the P450soy protein was called soyC. Five  
CC nucleotides downstream from the stop codon for soyC another ORF was  
CC identified. This ORF encodes an apparent ferredoxin-like protein.  
CC The gene was designated soyb and the protein ferredoxin-soy.  
SO Sequence 412 AA;

Query Match 14.5%; Score 316; DB 1; Length 412;







KM Daunomycin C-14 hydroxylase; doxa gene; doxorubicin; daunomycin;  
KW 13-dihydrocarminomycin; carminomycin; anthracycline;  
KW anticancer; cytostatic; cancer; therapy; plasmid pANT199.  
OS Chimeric - Streptomyces sp. strain C5.  
FH Chimeric - synthetic.  
FH Key location/Qualifiers  
FT Cleavage-site 31  
FT Protein /note="enterokinase cleavage site"  
FT 53..474  
PN MO974439-A2. /note="native daunomycin C-14 hydroxylase"  
PD 27-NOV-1997.  
PF 22-MAY-1997; 008690.  
PR 24-MAY-1996; US-653650.  
PA (OHIS ) UNIV OHIO STATE RES FOUND.  
PI Desanti CL, Dickens ML, Strohl WA;  
DR WPI: 98-018495/02.  
DR N-PSDB: V01452.  
PT Producing doxorubicin from daunomycin by daunomycin C14 hydroxylase  
PT - also hydroxylation and oxidation of other anthracycline(s) with  
PT the same enzyme  
PS Disclosure: Fig 9: 59pp; English.  
CC This protein comprises an N-terminal modified enzyme derived from  
CC the daunomycin C-14 hydroxylase (see W36128) of Streptomyces sp.  
CC strain C5. It is encoded by a gene construct (see V01452) in  
CC plasmid pANT199. In this plasmid, the doxa gene (see V01447) is  
CC translationally fused with a leader sequence encoding 6  
CC histidine residues so that the fusion protein can be affinity  
CC purified on a nickel-agarose gel. Daunomycin C-14 hydroxylase  
CC is a P450-like enzyme capable of converting daunomycin to the  
CC anticancer agent doxorubicin. Host cells, especially Streptomyces  
CC host cells, transformed with pANT199 can be used in methods for  
CC the production of doxorubicin from daunomycin or for the  
CC hydroxylation and oxidation of other anthracyclines.  
SQ Sequence 474 AA;

Query Match 7.4%; Score 161; DB 1; Length 474;  
Best Local Similarity 23.6%; Pred. No. 7.9e-08;  
Matches 96; Conservative 51; Mismatches 205; Indels 54; Gaps 14;

QY 39 QEAMAVYQESNVDDLVWTRNGGHMTITRGQLREATEDYRHSSE--CFPIREAGEAY 96  
DB 83 REAGPVV-EVNAP-----AGPRAWITDDALREVLADPRFVKGPDLPATAMRGVDDGL 135  
QY 97 D-----PIPTMDPPRQOFRALANQVGMVVNDKLENRIOELACSLIESL-----R 143  
DB 136 DIVPELRPTLLAVDGEDHRLRLRIHARAFNFRRLAERTDRIATAIDRLLELADSSDR 195  
QY 144 PQGQCNFTEDYAEFPPIRIFMLAGLP-----EEDIPHLKYL-----TDQMTRPD 188  
DB 196 SGEPAPLIGGFAYHFPILVLCGLGVPTDPAAREAVGVKALGLGPGSAGGDTDP 255  
QY 189 GSKTFPAEAKALDYLIPTIEQRQKGTDAISIVANGQVNGRPIITSDEAKRMCGLLVG 248  
DB 256 GDVPDTSALSL--LLEAVHAARRKDTRTMTRVLYERRAQEFGVSDDQLVYMITGLIFA 313  
QY 249 GLDPTVVFELFSMEFLAKSPENHROELIERPERIPACECELLRR-----FSLVADGRILTS 303  
DB 314 GHDTTGSFLGF--LAEVLAGRLADADGDAISRYEEALRHNPVYISL--WRPAAT 367  
QY 304 DYEFGVQLKKGDQILLPOMLSGLDERENACPMHVDFSROKVSHTTGGSHLCLGQHILA 363  
DB 368 EVVIRGVRLPRGARVLVDIGSTMTDGRNHDAHAFHPRDRSRRLRFGDGRHICIGBQLA 427  
QY 364 KRELIVTLKEMLTRIPDESTA-PGAQIQ--HKSGIVSGVQALPLVW 406  
DB 428 QLESRTMIGVLRSPQARLAVPEELRMCRCGAQTARLTDLR-VW 472

---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: October 4, 2000, 06:41:03 ; Search time 66.92 Seconds

(without alignments)  
103.693 Million cell updates/sec

Title: US-09-246-451-2

Perfect score: 2160  
Sequence: 1 TTTETQSNAFLAPLPHVPE.....IVSGVQALPLVWDPAITTKAV 414Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 164575 seqs, 16761186 residues

Total number of hits satisfying chosen parameters: 164575

Minimum DB seq length: 0  
Maximum DB seq length: 200000000Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

## Database :

Issued Patents\_AA:\*  
1: /cgnl\_7/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgnl\_7/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgnl\_7/ptodata/1/iaa/6\_COMB.pep:\*  
4: /cgnl\_7/ptodata/1/iaa/PCITUS\_COMB.pep:\*  
5: /cgnl\_7/ptodata/1/iaa/backfilest1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	363.5	16.7	416	3	US-09-320-878-18
2	340	15.6	403	5	5212296-9
3	330	15.1	406	5	5212296-6
4	316	14.5	412	1	US-08-102-863-11
5	316	14.5	412	4	PCR-US92-10885-11
6	271.5	12.5	419	3	US-09-335-409-8
7	161	7.4	422	2	US-09-096-982-5
8	161	7.4	422	2	US-08-653-650A-5
9	161	7.4	474	2	US-09-096-982-8
10	161	7.4	474	2	US-08-653-650A-8
11	158	7.2	443	2	US-09-096-982-9
12	158	7.2	443	2	US-08-653-650A-9
13	156	7.2	422	1	US-08-396-218-2
14	156	7.2	422	1	US-08-760-116-2
15	129	5.9	512	2	US-08-194-981E-5
16	123.5	5.7	512	2	US-09-320-878-7
17	118	5.4	513	3	US-08-948-564-6
18	115.5	5.3	490	1	US-08-201-118-7
19	115.5	5.3	490	4	US-08-238-821B-7
20	115.5	5.3	490	4	PCR-US95-05744-7
21	113.5	5.2	490	1	US-08-201-118-3
22	113.5	5.2	490	1	US-08-201-118-9
23	113.5	5.2	490	2	US-08-238-821B-3
24	113.5	5.2	490	2	US-08-238-821B-9
25	113.5	5.2	490	4	PCR-US95-05744-3
26	113.5	5.2	490	4	PCR-US95-05744-9
27	113	5.2	501	3	US-08-906-791-2
28	111.5	5.1	504	1	US-08-457-274A-25

29	111.5	5.1	504	4	PCR-US95-05758-25	Sequence 25, App1
30	110	5.0	492	3	US-08-724-466B-2	Sequence 2, App1
31	106.5	4.9	496	1	US-08-313-075A-50	Sequence 50, App1
32	102.5	4.7	472	2	US-08-622-166A-2	Sequence 2, App1
33	102.5	4.7	472	2	US-08-622-166A-4	Sequence 4, App1
34	102	4.7	497	3	US-08-724-466B-4	Sequence 4, App1
35	99.5	4.6	490	1	US-08-201-118-1	Sequence 1, App1
36	99.5	4.6	490	2	US-08-238-821B-1	Sequence 1, App1
37	99.5	4.6	490	4	PCR-US95-05744-1	Sequence 1, App1
38	99.5	4.6	498	1	US-08-457-274A-24	Sequence 24, App1
39	99.5	4.6	498	2	PCR-US95-05758-24	Sequence 24, App1
40	97.5	4.5	500	2	US-08-314-601-2	Sequence 2, App1
41	97.5	4.5	500	4	PCR-US95-13051-2	Sequence 2, App1
42	94.5	4.3	523	3	US-08-606-505B-67	Sequence 67, App1
43	94.5	4.3	1026	1	US-07-998-003A-95	Sequence 95, App1
44	94.5	4.3	1026	1	US-08-453-274B-95	Sequence 95, App1
45	94.5	4.3	1026	1	US-08-453-695A-95	Sequence 95, App1

## ALIGNMENTS

```
RESULT 1
US-09-320-878-18
; Sequence 18, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320, 878A
; EARLIER FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141, 908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073, 538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846, 247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119, 139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100, 880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087, 080
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-320-878-18
```

Query Match 16.7% Score 363.5; DB 3; Length 416;  
Best Local Similarity 28.3%; Pred. No. 1,3e-29;  
Matches 97; Conservative 61; Mismatches 164; Indels 21; Gaps 7;

```
QY 76 EDNRHFSSEPPFLIREGEAYDFIPMSDPPEDROPALANOVVGMVYKLEMRIOELA 135
DB 72 KDMR--NSTPFLDEALAHNMHIES--DPPHTRFLKLVAREFTMRVVELLRPROVEIV 127
QY 136 CSLIESL--RPGQCNTEDYAEFPPIRIPMLAGLEEDIPHLKXYLTDOMTSPDSMTF 193
DB 128 DGLVDALAPADGARDLMEFLAMPPLITVISELGVPEPDRAAFRWTDAFVPPDPAQA 167
QY 194 AEAKKALVDLPIITIEORRCKPPTDAIS--IVANGVNGRITISDEAKRKGGLLVGLDT 252
DB 188 QTMAEMSEGYLSRLIDSKRGQDEDDLSALVTSDEGSRUTSEELGNMHLIVAGHET 247
```

Thu Oct 5 09:02:40 2000

us-09-246-451-2.rai

Page 2

[illegible]

```

RESULT 2
5212296-9
PATENT NO. 5212296
APPLICANT: DEAN, CAROLINE; HARDER, PATRICIA A.; LETO, KENNETH
J.; O'KEEFE, DANIEL P.; OMER, CHARLES A.; ROMESSER, JAMES A.
TEPPERMAN, JAMES M.
TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING
CYTOCHROMES
NUMBER OF SEQUENCES: 19
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/569,781
FILING DATE: 23-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 464,499
FILING DATE: 12-JAN-1990
APPLICATION NUMBER: 405,605
FILING DATE: 11-SEP-1989
SEQ ID NO: 9
LENGTH: 403
5212296-9

```

Query Match 15.6%; Score 340; DB 5; Length 403;  
Best Local Similarity 28.9%; Pred. No. 3.5e-27;  
Matches 105; Conservative 63; Mismatches 165; Indels 30; Gaps 13

OY	63	WIATPGCOLIREAYEDYRHFSSEC-----PPIPREAGEAYDEIP--SMDPEQORQFALA	115
Db	52	WLATYRHQVRAVLGDPR-FSDAHRTGEPPLTAGRGELICTGNPFLMDDDEBARLRRL	110
OY	116	NOVGMAPVVDKLENRIOELACSLIESLP--OGOCNTEDEYAEPPRIIFMLGLPREDI	174
Db	111	TADFIYKKEVLEAMREPVGLADVIDYDKMTGTGSADIVTEFALPLPSLYTCLLGVPEDH	170
OY	175	PHL----KYLTQMTRPDGSMTFAEKALYDYLPIIEOROKPGTDAIS-IVANGVN	229
Db	171	AFEOERSVLLTLTRSTPE---EVRAAODELELEYLARLARTKREPRDDAITSRLVAERGE	227
OY	230	GRPTISDEAKMKCGLLLVGLDITYVNFLSFMEFLASPEIKRQLLIRPEIIPACCELL	289
Db	228	DTQIAT-----MGRLLVAGHETTANMTALSTLVLLRNPDQARLARAPALVKGAVEEL	282
OY	290	RRESLVADG--RIITSDEYFHYGVLKKQDOLLQOMLSGDREBNAC--MHYDFSRQV	345
Db	283	KYLTIYVHNGVRIATEDVLIGGRITLAGECYLC--MISSANKRAEVRPGGDLDIVADAR	340
OY	346	SHITTFHGSHLCLGQHLARREIYTLKEMTLRIPEDFSIA-PGAOIQKSGI-VSGVALP	403
Db	341	RHVAVGFEVQCLGQPLARVLEQLAIEFLRLRDLRLAYVHEIIPPRGDMATGVASHLP	400
OY	404	LWV 406	
Db	401	IAM 403	

RESULT 3  
5212296-6  
; Patent No. 5212296  
; APPLICANT: DEAN, CAROLINE; HARDER, PATRICIA A.; LETO, KENNETH  
; J.; O'KEEFE, DANIEL P.; OMER, CHARLES A.; ROMESSER, JAMES A.

```

TEPPERMAN, JAMES M.
; TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING
; CYTOCHROMES
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/569,781
; FILING DATE: 23-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 464,499
; FILING DATE: 12-JAN-1990
; APPLICATION NUMBER: 405,605
; FILING DATE: 11-SEP-1989
; SEQ ID NO:6:
; LENGTH: 406
; 5212296-6

```

Query Match	15.1%:	Score 330;	DB 5;	Length 406;
Best Local Similarity	26.3%;	Pred. No. 3.9e-26;		
Matches 104;	Conservative	74;	Mismatches 193;	Indels 24;
			Gaps	11.
OY	31	PSNLAGVO--EAMVILQESNPDLVWTRCNGH-WIATRCOLIREAYDYR-----	79	
		: : : : :   : : : :   : : : :   : : : :   : : : :		
Db	17	PSNRCPQVLPQGVAGQLRRTPEGRILRVTLYDQRQAMVYVYKHEARKLLDPPLSNRITD	76	
OY	80	HFSSECPITP-RENGEAVDLFPTSMDDPEQROFRALANOVYGMVYOKLENRIOELACS	137	
		: : : : :   : : : :   : : : :   : : : :   : : : :		
Db	77	NEPAPSREAVRESPQAF---IGLDPEHETRRMTISEFTVKIKIMREVEVEVWG	132	
OY	138	LIESLRPGQ-QCNFEDVAFPPRIPIFMLLAGLPREDIPIHLYTFQMPRPGSMFAEA	196	
		: : : : :   : : : :   : : : :   : : : :   : : : :		
Db	133	FIDEMLAGPTADLYSQFALPVPNSWICLLGVPAPADHEFPDQASRLVQSDAGSALTA	192	
OY	197	KEALYDLPIITEORRQKPGTDAI-SIYANGVNGRPITSDEAKRMGILLVGDLTPVN	255	
		: : : : :   : : : :   : : : :   : : : :   : : : :		
Db	193	RNDLIGYDLGITTCOTGTEAGLVCGLVADQLNGE-IDREELISTAMMLLAGHETTAS	251	
OY	256	FLTSFAMEFLAKSPHEHQELIERPERIPACCELLREFSL-VADGRILSDVEFHVOAK	313	
		: : : : :   : : : :   : : : :   : : : :   : : : :		
Db	252	MISLSVITLLDHPQYAAIRADRSLVPGCAVEELRLRLADYAGGVATADVEVBHLIR	311	
OY	314	KGDQILPQMLSGLDERENACPMHVDFSRQKSHTTFFGHSHLCLQHLARREIIVTKE	373	
		: : : : :   : : : :   : : : :   : : : :   : : : :		
Db	312	AGEGIYIVANSINRDGTVEYEDPDALDIHRSARHHLAFGFGVQCIGQNLARLEVLITLNA	371	
OY	374	WLTRIPDPSIA-PGAOIOHKSG-IYSQVADPLVW	406	
		: : : : :   : : : :   : : : :   : : : :   : : : :		
Db	372	IMDRVETRLAAVPEQLVLRPGTTIGGVNELPVTW	406	

```

1  RESULT 4
2  US-08-102-863-11
3  ; Sequence 11, Application US/08102863
4
5  Patent No. 546590
6
7  GENERAL INFORMATION:
8
9  APPLICANT: SAKIASLANI, SIMA
10
11 TITLE OF INVENTION: CONSTITUTIVE
12
13 TITLE OF INVENTION: EXPRESSION OF P45
14
15 TITLE OF INVENTION: AND FERRODIXIN-SO
16
17 TITLE OF INVENTION: STREPTOMYCES
18
19 NUMBER OF SEQUENCES: 12
20
21 CORRESPONDENCE ADDRESS:
22
23 ADDRESSEE: E. I. DU PONT DE NEMOURS
24
25 ADDRESSEE: AND COMPANY
26
27 STREET: 1007 MARKET STREET
28
29 CITY: WILMINGTON
30
31 STATE: DELAWARE
32
33 COUNTRY: USA
34
35 ZIP: 19898
36
37 COMPUTER READABLE FORM:
38
39 MEDIUM TYPE: Floppy disk
40
41 COMPUTER: IBM PC compatible
42
43 OPERATING SYSTEM: PC-DOS/MS-DOS

```



SEQ ID NO 8  
LENGTH: 419  
TYPE: PRT  
ORGANISM: Sorangium cellulosum  
US-09-335-409-8

Query Match 12.5%; Score 271.5; DB 3; Length 419;  
Best Local Similarity 23.8%; Pred. No. 5.2e-20;  
Matches 100; Conservative 67; Mismatches 179; Indels 75; Gaps 14;

QY 6 QSNANAPLPPEHIVPDEDMYNSLSAGVQ-----AMAVLOSNNPDLVWTRCNG 61  
DB 3 QEQANQSETRP-----AFDFKFPAP-----GYAEDFPFAIERLRBA-TPIFYWD--EGR 48  
QY 62 HWIATRGQ-----LIREAYEDYRHFSSECPFIIPREAGEAVDFIPTSMDPEQR 109  
DB 49 SWVLTIRHVSANFRDBERFVNSREWESSAEYSSAIP-----ELSDMKKIGLFGLPEDHA 104  
QY 110 QFRALANOVGMFVVDKLENRIQELACLIESLRPOGOCNFTEDVAPPEPIRIFMLLAGL 169  
DB 105 RVKLVNPSFTSRVIDLRLRAEIQRIVDQLLDANSQGEFVDVADYAEIGPMRAISLAKV 164  
QY 170 PEEDIHKLKLTQMTRRPGDSMTFAEKELDYLP-----I 207  
DB 165 PAF-----CDEKFRFSGSAT---ARALGYLVPOVDEETKTLVASYTEGLALLHDV 212  
QY 208 IEBRROKP-GTDAISIVANGVGRPITSDEAKRMCGLLLVGLDVIYVNFSLPSMEFLAK 266  
DB 213 LDDRKNPLNDVLTMLQAEADGSRSLKELVALVGAITATDTTYIATAVNLRLR 272  
QY 267 SPEHROELIERPERIPACCELLRFSVLADG--RILTSDEPHGVOLKKGOI--LLPQ 322  
DB 273 SPEALVLAEPGLMNAIDEVLRFDNILRLIGVRFARQDLERCGASIKKGEVFLIIPS 332  
QY 323 MSLGDERENACPMHVDFSRQKVSHTTFGHSILCLGQHLARREIIVTLKWLTRIPDES 362  
DB 333 ALR--DGVTFSRDVPDVRDGTGASLAYGRGPHVCPGVSILAEALAVGTIRRPPEMK 390  
QY 383 I 383  
DB 391 L 391

## RESULT 7

US-09-096-982-5  
Sequence 5, Application US/09096982  
Patent No. 5962293  
GENERAL INFORMATION:  
APPLICANT: Strohl, William R.  
APPLICANT: Dickens, Michael L.  
APPLICANT: Desanti, Charles L.  
TITLE OF INVENTION: METHOD OF PRODUCING DOXORUBICIN  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CALFEE, HALTER & GRISWOLD  
STREET: 800 Superior Avenue, Suite 1400  
CITY: Cleveland  
STATE: Ohio  
COUNTRY: USA  
ZIP: 44114-2688  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/096,982  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Goltick, Mary E.  
REGISTRATION NUMBER: 34829

REFERENCE/DOCKET NUMBER: 22727/00131  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 216-622-8458  
TELEFAX: 216-241-0816  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 422 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-096-982-5

Query Match 7.4%; Score 161; DB 2; Length 422;  
Best Local Similarity 23.6%; Pred. No. 1.8e-08;  
Matches 96; Conservative 51; Mismatches 205; Indels 54; Gaps 14;

QY 39 QEQAMAVLOSNNPDLVWTRCNGHWIATRGQILREAYEDYRHFSSE--CFPIPREAGEAV 96  
DB 31 REAGPVV-EVNAF-----AGCPAWITDDALAREVLADPRFVKGPDLPATMARGVDDGL 83  
QY 97 D-----FIPSMDBPQROFRALANOVGMFVVDKLENRIQELACLIESL-----R 143  
DB 84 DIPVPELRPTLIADVGEHRRRLRHADAFNPRRLAERTDRIATAIDRLTELADSSDR 143  
QY 144 PQGOCNFTEDVAPPEPIRIFMLLAGLP-----EEDIPHLKYL-----TDQMTRPD 188  
DB 144 SCGRAPLLIGFAFHPFLVLCGLGVYVDPMARARAVGLALGAGQSGAGSDGTDPA 203  
QY 189 GSWTFAEAKELAYDYLPIIEORRQKPGTDAISIVANGVNGRPITSDEAKRMCGLLLVG 248  
DB 204 GVDPTSALEST--LLEAVHAARKDTRMTRVLYERRAQEFQVSDDQLVYMITGLIPA 261  
QY 249 GLDVTYVNFSLFSMEFLAKSPENHRELIERPERIPACCELLR-----FSLVADGILTS 303  
DB 262 GHDTSFSGF--LLAEVLAGRLAADADGDAISRPEBALRHHPVPVSL--WTFPAAT 315  
QY 304 DYEFHGVOLKKGOILLPQMLSGDERENACPMHVDFSRQKVSHTTFGHSILCLGQHLA 363  
DB 316 EYVIRGVRLPRGAPVLDIEGTTDGRHHDAPHAIFPDRSRRLTLFGGPHYCTIGBOIA 375  
QY 364 RREIIVTLKWLTRIPDFSIA-PGAQIQ--HNSGIVGVALPLVW 406  
DB 376 QLESRTMIGVLRSPQARLAVPEYELRWCKRKAQYARLTDL-P-VW 420

## RESULT 8

US-08-653-650A-5  
Sequence 5, Application US/08653650A  
Patent No. 5976830  
GENERAL INFORMATION:  
APPLICANT: Strohl, William R.  
APPLICANT: Dickens, Michael L.  
APPLICANT: Desanti, Charles L.  
TITLE OF INVENTION: METHOD OF PRODUCING DOXORUBICIN  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CALFEE, HALTER & GRISWOLD  
STREET: 800 Superior Avenue, Suite 1400  
CITY: Cleveland  
STATE: Ohio  
COUNTRY: USA  
ZIP: 44114-2688  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/653,650A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:





APPLICATION NUMBER: US/08/653.650A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Golick, Mary E.  
REGISTRATION NUMBER: 34829  
REFERENCE/DOCKET NUMBER: 22727/00131  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 216-622-8458  
TELEFAX: 216-241-0816  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 474 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-653-650A-8

Query Match 7.4%; Score 161; DB 2; Length 474;  
Best Local Similarity 23.6%; Pred. No. 2.2e-08;  
Matches 96; Conservative 51; Mismatches 205; Indels 54; Gaps 14;

QY 39 OEAMAVLOESNPDLVWTRCNGHMATRGQLIREAYEDYRHFSS--CPFIPTREAGEAY 96  
DB 83 REAGPVV-EVNAP-----AGGPAAVITDDALAREVLADPRVKGPDLPATMKGVDGL 135  
QY 97 D-----FIFTSMDPREQROFRALANOVGMVVDKLENIQELACSLIESL-----R 143  
DB 136 DIPVELRPFLLIADGEGHRLRLRIHAFNPRRLAERTDRIADRLTELADSSR 195  
QY 144 POGQCNFTEDVAPERRIRIFMLAGLP-----BEDIPHLKYL-----TDQMTRPD 188  
DB 196 SEPAELLIGGFAYHNPFLVICELGVPTDPMAREAVGVALKALIGGPGSAGGDTDA 255  
QY 189 GSMTFAEAKALYDLIPIIEOROKPGTDAISIVANGVNGRPITSDAKRMCGILLVG 248  
DB 256 GVPPTSALESU--LLEAVHAARRKDTRTMTVLYERAAEGSVSDQLVYMITGLIFA 313  
QY 249 GUDTVVNFISFMEFLANSPEHROELIERPERIPACEELLRR-----PSLVADRGLTS 303  
DB 314 GHDITGSGFGF---LLAEVLAGRLAADADGDAISRFEVALRHHNPVPTSL--WRFAT 367  
QY 304 DYEHGVOLKKGDQILLPOMLSGLDERENACPMHYDFSROKVSHTTGHSHLCUGHLA 363  
DB 368 EVVINGVRLPGAPVLYDIEGTNTDGRHHADAPHAPRPSRRRLTFGDPHYCIGEQLA 427  
QY 364 RELIIVTLKEMLTIRIPDSIA-PGAQIO--HKSQIVSGVALPLW 406  
DB 428 QLESRTMIGVLSRFPQARLAVPYEELRWCKRGAQTARLTDLF-VW 472

RESULT 11  
US-09-096-982-9  
Sequence 9, Application US/09096982  
Patent No. 5962293  
GENERAL INFORMATION:  
APPLICANT: Strohl, William R.  
APPLICANT: Dickens, Michael L.  
APPLICANT: Desanti, Charles L.  
TITLE OF INVENTION: METHOD OF PRODUCING DOXORUBICIN  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CALFEI, HALTER & GRISWOLD  
STREET: 800 Superior Avenue, Suite 1400  
CITY: Cleveland  
STATE: Ohio  
COUNTRY: USA  
ZIP: 44114-2688  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/096.982  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Golick, Mary E.  
REGISTRATION NUMBER: 34829  
REFERENCE/DOCKET NUMBER: 22727/00131  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 216-622-8458  
TELEFAX: 216-241-0816  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 443 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-096-982-9

Query Match 7.2%; Score 158; DB 2; Length 443;  
Best Local Similarity 23.6%; Pred. No. 4.1e-08;  
Matches 95; Conservative 50; Mismatches 210; Indels 48; Gaps 13;

QY 39 OEAMAVLOESNPDLVWTRCNGHMATRGQLIREAYEDYRHFSS--CPFIPTREAGEAY 96  
DB 52 REAGPVV-EVNAP-----AGGPAAVITDDALAREVLADPRVKGPDLPATMKGVDGL 104  
QY 97 D-----FIFTSMDPREQROFRALANOVGMVVDKLENIQELACSLIESL-----R 143  
DB 105 DIPVELRPFLLIADGEGHRLRLRIHAFNPRRLAERTDRIADRLTELADSSR 164  
QY 144 POGQCNFTEDVAPERRIRIFMLAGLP-----BEDIPHLKYL-----TDQMTRPD 188  
DB 165 SEPAELLIGGFAYHNPFLVICELGVPTDPMAREAVGVALKALIGGPGSAGGDTDA 224  
QY 189 GSMTFAEAKALYDLIPIIEOROKPGTDAISIVANGVNGRPITSDAKRMCGILLVG 248  
DB 225 GVPPTSALESU--LLEAVHAARRKDTRTMTVLYERAAEGSVSDQLVYMITGLIFA 282  
QY 249 GUDTVVNFISFMEFLANSPEHROELIERPERIPACEELLRRFSV--ADGRILTSDE 306  
DB 283 GHDITGSGFGF---LLAEVLAGRLAADADGDAISRFEVALRHHNPVPTSLWRFATEV 339  
QY 307 FHGVOLKKGDQILLPOMLSGLDERENACPMHYDFSROKVSHTTGHSHLCUGHLARE 366  
DB 340 IRGVRLPGAPVLYDIEGTNTDGRHHADAPHAPRPSRRRLTFGDPHYCIGEQLA 399  
QY 367 ILVTLKEMLTIRIPDSIA-PGAQIO--HKSQIVSGVALPLW 406  
DB 400 SRTMIGVLSRFPQARLAVPYEELRWCKRGAQTARLTDLF-VW 441

RESULT 12  
US-08-653-650A-9  
Sequence 9, Application US/08653650A  
Patent No. 5976830  
GENERAL INFORMATION:  
APPLICANT: Strohl, William R.  
APPLICANT: Dickens, Michael L.  
APPLICANT: Desanti, Charles L.  
TITLE OF INVENTION: METHOD OF PRODUCING DOXORUBICIN  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CALFEI, HALTER & GRISWOLD  
STREET: 800 Superior Avenue, Suite 1400  
CITY: Cleveland  
STATE: Ohio  
COUNTRY: USA  
ZIP: 44114-2688  
COMPUTER READABLE FORM:

```

? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/653,650A
? FILING DATE:
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: Golrick, Mary E.
? REGISTRATION NUMBER: 34829
? REFERENCE/DOCKET NUMBER: 22727/00131
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 216-622-8458
? TELEFAX: 216-241-0816
? INFORMATION FOR SEQ ID NO: 9:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 443 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: peptide
? OS-08-653-650A-9

```

Query Match	7.2%	Score 158	DB 2	Length 443
Best Local Similarity	23.6%	Pred. No.	4.1e-08	
Matches 95	Conservative 50	Mismatches 210	Indels 48	Gaps 13

QY	39	QEAANAIVLOESNVPLVYTRCNGXMHITRQGLIEAEYEDXKHFSS--CFPIPEAGEAY	96
Db	52	REAGPV- EVNAP-----ACGPAWVITTDALAEVLADRFVYVDPDLAPAMKGVYDGL	104
QY	97	D-----FIPSTMDEPREROFBALANQVGMFVDKLENRIQELACSLIESL-----R	143
Db	105	DIPVPELRFPLLLVADGSDHNRRLRIRIHAFAFNPRAERLERTDRITAIADRLITELADSSDR	164
QY	144	PGGCGNFEDYAEFPPIRIFMLLAGLP-----EEDIPHLKYL-----TDQMTRPD	188
Db	165	SGEPAELIGGFAHYHPRPLVYICELGVRVTDPRAMAREVAVGLKALGLGPGSAGGSDGTDPA	224
QY	189	GSMFAEAKKALVYLYLRPIEDRQKQGTDAISLVANGQVNGRITGDEAKRMQGLIYVC	248
Db	225	GDVDPDTSLISL--LLAEVAHAAKKDRTRTMYLVERAQAQFSGVSODQLVMTTGIIFA	283
QY	249	GIDTVNVLISFSEMEFLAKSPREHQELTEREPRIACEELLRRFSY--ADGRITGSDYE	306
Db	283	GHDTTGSFLGF--LLAEVLALGRILAADGDALSRFVEALRNHRPPRYLLMFRATEVY	339
QY	307	PHGVOLAKKDDOILLPOMLSGLDERENACPMHWFSROKVSHTTGFHGSHLACLOHARE	366
Db	340	IRGVRLPRGAPVLVDIGTINTDGRHNHDAPAHFPRDRFSRRRLTGDDGRPHVICIGQLANLE	399
QY	367	IIVYLKEMLRIFIPDSTA-PCAQIQ--HKSGISGVQALPLVW	406
Db	400	SRIMIGVLRSRFPQARLAVPVEELRMCRKGAQTARILDDL--VW	441

RESULT 13  
US-08-396-218-2  
Sequence 2, Application US/08396218  
Patent No. 5695966  
GENERAL INFORMATION:  
APPLICANT: INVENTI, Augusto  
APPLICANT: BREME, Umberto  
APPLICANT: COLOMBO, Anna L  
APPLICANT: HUTCHINSON, Charles R  
APPLICANT: OTTEN, Sharee  
APPLICANT: SCOTTI, Claudio  
TITLE OF INVENTION: PROCESS FOR PREPARING DOXORUBICIN  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIKADO, MARCELSTEIN, MURRAY & ORAM

STREET: 655 Fifteenth Street, N. W., Suite 330 - G  
STREET: Street Lobby  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-5701  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/396,218  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: KITTS, Monica C  
REGISTRATION NUMBER: 36,105  
REFERENCE/DOCKET NUMBER: PI615-5002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202/638-5000  
TELEFAX: 202/638-4810  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 422 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-396-218-2

Query Match 7.28; Score 156; DB 1; Length 422;

[illegible]

RESULT 14  
 US-08-760-116-2  
 ; Sequence 2, Application US/08760116  
 ; Patent No. 5766150  
 ;  
 ; GENERAL INFORMATION:  
 ;  
 APPLICANT: INVENTI, Augusto  
 APPLICANT: BREME, Umberto  
 APPLICANT: COLOMBO, Anna L.  
 APPLICANT: HUTCHINSON, Charles R.

APPLICANT: OTTEN, Sharee  
 APPLICANT: SCOTTI, Claudio  
 TITLE OF INVENTION: PROCESS FOR PREPARING DOXORUBICIN  
 NUMBER OF SEQUENCES: 4  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: MIKAIKO, MARCELSTEIN, MURRAY & ORAM LLP  
 STREET: 655 Fifteenth Street, N. W., Suite 330 - G  
 STREET: Street Lobby  
 CITY: Washington  
 STATE: DC  
 COUNTRY: USA  
 ZIP: 20005-5701  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/760,116  
 FILING DATE: 3-DEC-1996  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/396,218  
 FILING DATE: 27-FEB-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: KUTTS, Monica C  
 REGISTRATION NUMBER: 36,105  
 REFERENCE/DOCKET NUMBER: PI615-6007  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202/638-5000  
 TELEFAX: 202/638-4810  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 422 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 OS-08-760-116-2

	Query Match	7.2%;	Score 156;	DB 1:	Length 422;	
	Best Local Similarity	23.8%;	Pred. No. 6.1e-08;			
	Matches	97;	Conservative	51;	Mismatches 202;	Indels 58; Gaps 15
OY	39	QEMAWAIOESNVVDLWTRCGNHVIATRGQIIRAEAYEDYRHFSSCEPIIPREAGEAYDF	98			
Dd	31	REAGFVV-EVNAP-----AGGPAVITDDALAREVLADPFEVKD-----PDLAAPAMWG	78			
OY	99	IPTSMND--PRROFORALA-----NQVGMRYVD-----KLENRIDELACSLIESL-	142			
Dd	79	VDDGDIDIVPELRPTLLAVDGSEAHRRRLRNAPARNPRRLAKERTDYIAIGRLITEIA	138			
OY	143	----RPQGOCNFTEDEYAERFPRIKIMLAGLP-----EEDIPIHLYL-----TDQ	183			
Dd	139	DASGRSGRAPELLGGFAVHPFLVICELGLGVTVTDPRAMAREAVSYLKALGIGGPOSGGD	198			
OY	184	MTRPRGSMTEFAKALKLDYLRTIEQRQAKGSTAISLVANGQYNAGRITSDEAKRMGC	243			
Dd	199	GTPRAGGVBDTSALSLSL--LLEAVSARKNDRPTMTRVLYERAQAQEFGSVSDOLUYMIT	256			
OY	244	LILVGLGLTJVVFELSFSMEFLAKSPENHRELIERPERIPAACEELLRRSLV--ADGRLL	301			
Dd	257	GLIFRGNHTTGSFLCF---LLAEVLAGRLAADADEDAVSRYFEALRIPIRPYYTIIMRA	313			
OY	302	TSDYEFGHVQALKKGOILLPOMLSGLDERENAKPMHWDFSROKVSHTTGHSHTLGION	361			
Dd	314	ALEVTLIGSVRLRPGARVLVDICTMTOGNGNHDARYAFHFDREBSWRMLRGCDRNYCTGQ	373			
OY	362	LARRLIYLTLEMELTRIIPDFSLA-PGAOIQ--HKSGIYSVALAPLW	406			
Dd	374	LAQESRTMIGVLRSRFEPEARLAVPYUDELRMCRKCAQOTARILELP-VW	420			

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

	Query Match	5.9%	Score 129;	DB 2:	Length 512;	
	Best Local Similarity	22.4%;	Pred. No. 5.4e-05;			
	Matches	95;	Conservative	67;	Mismatches	156;
					Indels	104;
					Gaps	22.
QY	27	DMYNSNLSAGVOEAMVALOESNPDLVWTRCGNGHMITRQLIREAVEDYRH-----	81			
		:   :   :   :				
Db	107	DLTYTTLISNQSMFS-----PD-----SGVMAAR--RLMONGKSTSIAD	149			
QY	82	--SSECFIPREAGEADVIFPTSD---PPEOROFR---ALANOVGMFVDKLENR	130			
		: :   :   :   :   :		:   :   :   :	:   :	
Db	150	PASSSCYLEHEVHSKEAEVLITLDELMAFGHPFNRYVVVSYNVICAFICGRYDHN	209			
		:   :   :   :   :		:   :   :   :	:   :	
QY	131	IQELACSLIESIRPGQCNEFTEDYAEPPIRIKFMILLAGIREDIDHLKYLTDPQMRPDS	190			
Db	210	HQEL-TLSIV-----NLNNFE-----VVGSSNPAEFILRYL-----PNPS	246			
QY	191	M-TFEAKREALDYLIPIIEORQ-----KPGTDASIYANGQ-----VNGRPITSDE	237			
		:   :   :   :   :				
Db	247	LNAFDLNEKFYSFKQKWKEHYKTFPEKGHLIDTD--SLIEHCQKOLDENAVQLSDE	304			
		:   :   :   :   :				
QY	238	AKRMGGLLVG-GLDTVVNFLSFSEMFIAKSPEHRQELIE-----RP-----ERI	281			
				:   :   :   :		
Db	305	KIINIVDDLFGFGPYTAITSMSLMYLVMNPRVQRKIQEELDLYIGSRRRRLDRSHL	364			
				:   :   :   :		
QY	282	P---AACELLLRRESLV--ADGRILITSDY EFPHGVOLKKGDQILLPOMLSGLDERENAC-	335			

Db 365 PYMEAFLETFRRHSSEVPFTIPHSTRTDTSIKGFYIPKGRCEVFNQMOIINHDKLWVNS 424  
QY 336 -----MHVDFSRQKV---SHTTFGHSHLCLGQHLARREIIVTLKKEWLTFRIPDFSIAP 385  
Db 425 EFLPERELTPDGAIDKVLSEKVIITFGMKRKCIGETVARWEVFLFLAILLQRV-EFSYPL 483  
QY 386 GAQI 389  
Db 484 GVKV 487

Search completed: October 4, 2000, 12:38:03  
Job time: 21420 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 05:49:35 ; Search time 73.09 seconds  
(without alignments)  
350.531 Million cell updates/sec

Title: US-09-246-451-2  
Perfect score: 2180  
Sequence: 1 TTERIQSNANLAPLEPHVPE.....IVSGVQALPLVWDPATRKAV 414

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 178050 seqs, 61884766 residues

Total number of hits satisfying chosen parameters: 178050

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_64:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2180	100.0	415	1	campor 5-monooxyg
2	380	17.4	393	1	cytochrome P450 Rv
3	371	17.0	410	2	cytochrome P450meg
4	365	16.7	397	1	cytochrome P450 my
5	362	16.6	396	1	cytochrome P450 yj
6	360	16.5	398	1	cytochrome P450 Rv
7	346.5	15.9	404	2	cytochrome P450 -
8	340	15.6	403	2	cytochrome P450 10
9	334.5	15.3	410	1	cytochrome P450 10
10	334.5	15.3	428	1	cytochrome P450 Rv
11	333	15.3	410	1	cytochrome P450 cy
12	331.5	15.2	405	1	cytochrome P450 cv
13	330	15.1	406	2	cytochrome P450 10
14	329	15.1	411	1	cytochrome P450 10
15	326	15.0	406	1	polyketide synthas
16	324.5	14.9	376	1	linalool 8-monooxy
17	320.5	14.7	405	1	polyketide hydroxy
18	318.5	14.6	395	1	cytochrome P450 Rv
19	318	14.6	310	2	cytochrome P450 bl
20	318	14.6	428	1	cytochrome P450 tl
21	316.5	14.5	399	1	cytochrome P450ter
22	316	14.5	412	2	cytochrome P450 -
23	313.5	14.4	410	2	cytochrome P450 (s
24	304	13.9	417	2	cytochrome P450 sc
25	304	13.9	438	1	cytochrome P450 ty
26	303	13.9	402	2	cytochrome P450 Rv
27	303	13.9	404	1	nitric-oxide reduc
28	296.5	13.6	408	2	cytochrome P450 no
29	296.5	13.6	408	2	cytochrome P450 no

30	294.5	13.5	406	1	S18531	cytochrome P450 er
31	293.5	13.5	337	2	T30601	cytochrome P450 hv
32	291.5	13.4	433	1	B70677	cytochrome P450 Rv
33	285	13.1	381	1	S15809	cytochrome P450 Cy
34	276.5	12.7	310	2	T44857	probable hydroxyla
35	266	12.2	412	1	B40634	erythromycin monoo
36	265.5	12.2	411	2	T36526	probable cytochrom
37	265	12.2	414	2	E70708	cytochrome P450 Rv
38	261.5	12.0	386	2	T30235	cytochrome P450 -
39	254	11.7	401	1	I40208	cytochrome P450 BJ
40	254	11.7	406	3	T17487	cytochrome P450 hv
41	248	11.4	411	1	JC4403	cytochrome P450 -
42	240	11.0	433	2	T44587	cytochrome P450 ho
43	236	10.8	402	1	A70707	cytochrome P450 Rv
44	235	10.8	396	1	H70730	cytochrome P450 Rv
45	225	10.3	391	3	T17486	cytochrome P450 hv

## ALIGNMENTS

RESULT 1  
O4PSCP  
campor 5-monooxygenase (EC 1.14.15.1) cytochrome P450 101 - Pseudomonas putida plasm  
N:Alternate names: cytochrome P450-CAM  
C:Species: Pseudomonas putida  
C:Date: 30-Apr-1982 #sequence.revision 31-Dec-1993 #text\_change 03-Mar-2000  
C:Accession: A25660; S34614; C60886; A00194  
R:Unger, B.P.; Gunsalus, I.C.; Sliagar, S.G.  
J. Biol. Chem. 261, 1158-1163, 1986  
A:Title: Nucleotide sequence of the Pseudomonas putida cytochrome P-450-cam gene and  
A:Reference number: A94678; MIMD:86111751  
A:Accession: A25660  
A:Molecule type: DNA  
A:Residues: 1-415 <UNG>  
A:Cross-references: GB:M12546; MID:g151114; PIDN:AAA25760.1; PID:g151115  
R:Atamaki, H.; Koga, H.; Sagara, Y.; Hosoi, M.; Horiochi, T.  
Biochim. Biophys. Acta 1174, 91-94, 1993  
A:Title: Complete nucleotide sequence of the 5-exo-hydroxycampor dehydrogenase gene  
A:Reference number: S34613; MIMD:93326643  
A:Accession: S34614  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-42 <ARA>  
A:Experimental source: ppg1, ATCC 17453; CAM plasmid  
R:Romeo, C.; Moriwaki, N.; Yasunobu, K.T.; Shastri, B.A.; Gunsalus, I.C.; Koga, H.  
J. Protein Chem. 6, 253-261, 1987  
A:Title: Identification of the coding region for the putidaredoxin reductase gene fro  
A:Reference number: A60886  
A:Accession: C60886  
A:Molecule type: DNA  
A:Residues: 408-415 <ROM>  
R:Handu, M.; Ames, L.G.; Yasunobu, K.T.; Shastri, B.A.; Gunsalus, I.C.  
J. Biol. Chem. 257, 42654-42671, 1982  
A:Title: Amino acid sequence of the Pseudomonas putida cytochrome P-450. II. Cyanogen  
A:Reference number: A00194; MIMD:83030788  
A:Accession: A00194  
A:Molecule type: protein  
A:Residues: 2-55, 58-276, 'Q', 278-361, 'S', 363-407, 'N', 409-415 <HAN>  
C:Genetics:  
A:Gene: camC; CYP101  
A:Genome: Plasmid  
C:Function:  
A:Description: catalyzes hydroxylation of campor to yield 5-exo-hydroxycampor; elec  
C:Superfamily: Pseudomonas plasmid campor 5-monooxygenase; cytochrome P450 homology  
C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygena  
F:246-380/Domain: cytochrome P450 homology <CYP>  
F:358/Binding site: heme iron (Cys) (axial ligand) #status experimental

Query Match 100.0%; Score 2180; DB 1; Length 415;  
Best local similarity 100.0%; Pred. No. 7, 6e-159;  
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy	1	TTEITQSNANLAPLPHPNPHNLVDFEDMNPNSLSAGIQEAMVALQESNVDLWTXKNG	60
Db	2	TTEITQSNANLAPLPHPNPHNLVEFEDMNPNSLSAGIQEAMVALQESNVDLWTXKNG	61
Oy	61	GHWIATROQLREAEVDYRHSSSECPFTPRAGEAYDPTPSMPPPEOROPRALANOVC	120
Db	62	GHWIATROQLREAEVDYRHHSSSECPFTPRAGEAYDPTPSMDPPEOROPRALANOVC	121
Oy	121	MPVVDKLENRIQELACSLIESLRPOGCNFTEDEYAEPPIRIFMLAGLPREDIPHLKYL	180
Db	122	MPVVDKLENRIQELACSLIESLRPOGCNFTEDEYAEPPIRIFMLAGLPREDIPHLKYL	181
Oy	181	TDQMTRPGCSMTFAKAKALDYLLPTTEOROKRGTAISIVANGOVNRPITSDEKR	240
Db	182	TDQMTRPGCSMTFAKAKALDYLLPTTEOROKRGTAISIVANGOVNRPITSDEKR	241
Oy	241	MCGLLLVGGLTVVNFLEFSMEFLAKSPENHOELLERPERIPACEBELLRRFSLVADCR	300
Db	242	MCGLLLVGGLTVVNFLEFSMEFLAKSPENHOELLERPERIPACEBELLRRFSLVADCR	301
Oy	301	LTSDFEFHGVOLKKGDQILLPOMLSGLDERBNACPMMHVDFSROKVSHTFEGHSGLCIG	360
Db	302	LTSDFEFHGVOLKKGDQILLPOMLSGLDERBNACPMMHVDFSROKVSHTFEGHSGLCIG	361
Oy	361	HLARREIIVTLKEWLTIRPDFSIABGAOIOHKSGIVSCVOLPLVMDBATTKAV	414
Db	362	HLARREIIVTLKEWLTIRPDFSIABGAOIOHKSGIVSCVOLPLVMDBATTKAV	415

RESULT 2  
C70929  
cytochrome P450 RV1785c - Mycobacterium tuberculosis (strain H37Rv)  
N:Contains: oxidoreductase (EC 1.-.-.-)  
C:Species: Mycobacterium tuberculosis  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: C70929  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajadream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987  
A:Accession: C70929  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-393 <COL>  
A:Cross-references: GB:A022021; GB:AL123456; NID:93250699; PIDN:CA17707.1; PID:e125459  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: RV1785c  
;Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology  
;Keywords: oxidoreductase

Query Match	17.4%;	Score 380;	DB 1;	length 393;
Best Local Similarity	28.7%;	Pred. No. 1.5e-21;		
Matches 110;	Conservative 71;	Mismatches 182;	Indels 20;	Gaps 10;

[illegible][illegible]

RESULT 3  
S39924  
Cytochrome P450meg - *Bacillus megaterium* (ATCC 13368)  
C:Species: *Bacillus megaterium*  
A:Variety: ATCC 13368  
C:Date: 07-Oct-1994 #sequence\_revision 23-Feb-1996 #excl\_change 04-Mar-2000  
C:Accession: S39924 S32216  
R:Rauschenbach, R.; Isernhagen, M.; Noeske-Junghut, C.; Boidol, W.; Siewert, G.  
Mol. Gen. Genet. 241, 170-176, 1993  
A:Title: Cloning sequencing and expression of the gene for cytochrome P450meg, the st  
A:Reference number: S39924; MUID:94049677  
A:Accession: S39924  
A:Molecule type: DNA  
A:Residues: 1-410 <RAU>  
A:Cross-references: EMBL:Z21972; NID:g288298; PIDN:CAA79985.1; PID:g288300  
A:Experimental source: ATCC 13368  
C:Superfamily: *Bacillus* cytochrome P450 CYP106; cytochrome P450 homology  
C:Keywords: chromoprotein; heme; iron; metalloprotein  
C:355/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match	17.0%;	Score 371;	DB 2;	Length 410;
Best Local Similarity	30.18;	Pred. No. 7.9e-21;		
Matches 109;	Conservative 53;	Mismatches 180;	Indels 20;	Gaps 7;

QY	63	WIATRGOLIRAYADYKHFSS-----CPIPREAGEAYDFIP-TSMDDPEOROPALA	115
Db	44	WNVKEYEDVKVYLSDYHNFSSVRKRTTISVGTDSSEGSVEPKIQITSESDPDHRRKSTLL	103
QY	116	NOVGMPEVVDKLENRIQELACSLIESLRPOGOCNFEDYAEPPIRFTMLAGPREDIP	175
Db	104	AAATPSRLQWMERIRIDEINDELIGQMDGCEIDIVASLASPLRTIYMALDKVPSKDL	163
QY	176	HLKYLTPDMTRPDGSMTPAE-----AKEALDYLIPIIEQRKRFGTAISIVANGQV	228
Db	164	LEKKWVDLLEFPREKREKOEYDKLQVAAREYUQYLYPIVQKRLNPADDIISDLKSEV	223
QY	229	NGRPTTSEAKRMGCLLVGSDIVVNFLSFM-EFLAKSPENQELIERPEIIPAACEE	287
Db	224	DGEMPTDDEVVRTMILLGACVETTSHLANSFYSLLYDDKEVYOELEHMLDIPQAVEE	283
QY	288	LIR-RESVLADGRITSDYEFHGVQLKKQDILLPQMLSGIDRENAACPMHVDFSR-QKV	345
Db	284	MLRRFENLIKIDRIYKEDNDLSTGLKEGSGSVVYVMSAAMDEMEFDPFTLIINHPRNK	343
QY	346	SHHTFGHSHLCGLQHLARRELIYTLKEWLTIRIDFSIABGAQI0HK--SGIVSGVAL	402
Db	344	KHLTFGNPHTCSGLPRLARLEAKIALALPAFLKKFKHLEAVPSQLEMLTDSATGQTLTSL	403
QY	403	PL 404	
Db	404	PL 405	

cytochrome P450 mycG - Micromonospora griseorubida



N:Contains: oxidoreductase (EC 1.-.-.-)  
C:Species: Micromonospora griseorubida  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: S51594  
R:Kinuye, M.; Takada, Y.; Muto, N.; Beppu, T.; Horinouchi, S.  
Mol. Gen. Genet. 245, 456-464, 1994  
A:Title: Characterization and expression of a P-450-like mycinamicin biosynthesis gene u  
A:Reference number: S51593; MUID:95107242  
A:Accession: S51594  
A:Molecule type: DNA  
A:Residues: 1-397 <IMO>  
A:Cross-references: EMBL:D16098; NID:g286050; PIDN:BA03672.1; PID:d1004187; PID:g303644  
C:Genetics:  
A:Gene: mycG  
A:Start codon: GTG  
C:Superfamily: Bacillus cytochrome P450 CYP106, cytochrome P450 homology  
C:Keywords: heme; oxidoreductase  
C:231-368/Domain: cytochrome P450 homology <CYP>

Query Match	16.7%;	Score 365;	DB 1;	Length 397;
Best Local Similarity	28.5%;	Pred. NO. 2.2e-20;		
Matches 105;	Conservative 57;	Mismatches 168;	Indels 38;	Gaps 8;

```

Qy      63 WIAITGQLIREAYEDYRHFSSCEPFI-----PREAGNAVDFIPTSDMPEDGQFRA 113
      | : || ||| | : : || | ||||| : |
Db      44 WLVTIR-----YEDYRAVLGDGRFVAGPSMTREDPRTRPEWVGGLLSDMPPEHSRLRR 96

```

```

Oy      114 LANOVGMPPVVDKLENRIOELACSLIESLRPOGO-CNFTEDYAEPPRIRIFMLAGLPEE 172
      | : | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      97 LVYAFARAEASLRPRAREIAHELVDQMAATGQPADLVAMFAROLPVRICELLCVPSA 156

```

```

QY 173 DIPHLKYLQDMTRPDGSM-----TFAEKALDYLIPIEQRRKPGTDAISIV 223
      | : | | : | | | : : | | : : | :
Db 157 D-----HDFRTNRSGAFLSTAEVTAEMQEAQAAYVMGDLDRRRKEPTDLSAL 209

```

224 ANGVNGRPTSDKRMCGLLVGGDVTVNFLSFMSMELAKSPHROELIERPERIPA 283  
 : : | : ||| : : : | | : : ||| | :  
 210 VQARDQDSISEQELDLAIGLVAGYESTTQIADFYVLLMTRELRLQDLDRPELIPS 269

QY 284 ACEEELRRFSL--VADGRILTSDYEHNGVLKKDQDILLPQLMSGLEDEENACRNHNVDF 340  
| | | | | | | : : : :  
Db 270 AVEETTRWVPLGVGTAFPRRYAVEDVTLRGVTIRAGEPVLASTGANRDAQRFPPADRIDV 329

Db 330 DRTPNOLTGSGHGVNHCLGAPLARVELQVALEVLLQRLPGIRGLGIPETQLRWSEGMILRG 389

QY	399	VQALPELVW	406
		: !	
Db	390	PLELPVW	397

RESULT 5  
B69051  
cvtrochrome p450 v11R - Bacillus subtilis

C:Accession: B69851  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Species: *Bacillus subtilis*  
C:Contains: Carboxylesterase (EC 3.1.1.1)

C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Car A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C. *Nature* 390, 249-256, 1997

Iech, J.; Harwood, C.R.; Henault, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masu

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato  
A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekow  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato,

A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis* 168  
A:Reference number: A69580; MUID:98044033  
A:Accession: B69851  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-396 <KUN>  
A:Cross-references: GB:J29110; GB:AL009126; NID:g2633472; PIDN:CAB13078.1; PID:el18322  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: yjiB  
C:Superfamily: *Bacillus* cytochrome P450 CYP106; cytochrome P450 homology  
C:Keywords: oxidoreductase  
E:236-371/Domain: cytochrome P450 homology <CYP>

Query Match	16.6%	Score 362	DB 1	Length 396
Best Local Similarity	30.18%	Pred. No. 3.6e-20		
Matches 99, Conservative	55	Mismatches 157	Indels 18	Gaps 7

OY    71 IREAYEDVRHSSCEPFIPEAGAYDFTPTSMPPEDQOFRALANOVGMPAVDKLEIR 130  
    ::: ||| |::|| ::|||::|::|||::|::|||::|::|||::|::|||::|::|||::|::|||::|::|||::|:  
Db 56 VKKVVGDDELFFS-C-MPOQTSGNST-INDDPKHTKIRSVVNKAFFTPIRWKKOEPR 111

---

189

Qy 131 IOELACSIIESLRGOGCFTEDAEPPFIRIFMLLAGIPEEDIPLHXYLTDOOT----- 185  
||| : || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 112 IQETDELQKFGORSEFDLVHDSYPLEVIVISELGLPSAHNEQGFAMSDLLVSTPKD 171

Oy 186 -RDGSMFAE---AKELALYDYLPILEORROKPGTASISVANGQVCNRPITSDEMR 240  
          |       |       |       |       |       |       |       |       |  
Db 172 KSEAEKAFLEERKCCEELAAFFAGLIEEKRNKPEDDIISLVEAEEGCKLSGEELLIP 231

QY 241 MCGLLVAGGLDIVNLFSTSMELFARKSPHROELETIERPRIIPACEELLR-RSLVAAGR 299  
| | | | | : : : : : |  
Db 232 ECFLLVAGNETTTNLISNAMYSLTETPGVYELLRSHPELMLQAVEELRRFRAPAVLRR 291

QY 300 LLSVIERHGVNLKGGDILLPLPQLSGLEIRENCAAPHNDFSKRVSHTEFGHSGLCLG 359  
 :  
 Db 292 IAKRDTETIGCHLIEGDWVLAFVASANRDEAKFDRPHMIDIRRHNPRIIAFGHGIIHFCLG 351

```

07 300 CHANRELVLVLENDMLKIFDE---SLAP 363
    ||| | : | : |
    ||| | : | : |
352 APLARLEANIATLSLSAFPHMECVSITP 380

```

RESULT 6  
H70807  
Cytochrome P450 RV3518C - Mycobacterium tuberculosis (strain H37Rv)

M: Genbank: *Salmonella* (acc 1:1:1)  
 C: Species: *Mycobacterium tuberculosis*  
 C: Date: 10-Sep-1999 #sequence\_revision  
 C: Accession: H70807  
 10-Sep-1999 #text\_change 10-Sep-1999

; Connor, R., Davies, R., Devlin, K., Fellwell, T., Gentles, S., Hamlin, N., Ho  
; Rajandream, M.A.; Rogers, J., Rutter, S., Seeger, K., Skelton, S., Squares, S.  
Nature 393, 537-544, 1998

A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complex  
A:Reference number: A70500; MUID:98295987  
A:Accession: H70807

A: Molecule type: DNA  
A: Residues: 1-398 <COL>  
A: Cross-references: GB:AL022022; GB:AL123456; NID:g3361554; PIDN:CAA17755.1; PIR:17755.1

C:Genetics: A:Gene: RV3518C  
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology

Query Match: 16.58; Score 360; DB 1; Length 398;

Matches	107;	Conservative	69;	Mismatches	169;	Indels	34;	Gaps	8;
QY	55	WTRCN-----GGIMATRGQLIREATEDYRHSS-----ECPIPRGEAYDIPT	101						
Db	26	WMRAQPVFRRDRNGLAASAYQAVDAIDAREPDLFNAGGIRPDQALP-----MMI	76						
QY	102	SMDDPEQGFQRLANQVGMPIVDKLENRIQELAGSLIESLRPGOCNFTEDYAEPPPR	161						
Db	77	DMDDPAHLIRKRLVAGFTTRKRVKDEASIALCPTLLIDAVCERECDVRDLAPLPMA	136						
QY	162	IFMLAGL-PEEDIPHLK-----YLTDMTRPDGSMTFEAKKALDYLLPIIPIQRQ	213						
Db	137	VIGDMLGVRPEGRDMELRMSDDLVTFLSHVSGEDFQITM-DAPFAYNDFPATIITAARRA	195						
QY	214	KRGDAISIVANGVNGRPITSDEAKRMCGLLVSGDLTVNVLFSFMEFLAKSEPHROE	273						
Db	196	PTDLDLVAVSVSEVDGERSLSDDELLVMEFTLLLTIGGDETTRTLTSGTEQLLRNDQDML	255						
QY	274	LIERPERIPACEELLRRSLVAD-GRITTSDEYERHGVLKKGDDIILPQMLSGUDEREN	332						
Db	256	LORDPSLLPGALIEELMRTAPYKNCRCRVLTADTERHGIALCAGEKMMILFESANDEAVE	315						
QY	333	ACPMHVDERSROKVSHTTGGHSGHLCLGOHLARREITIVLKLEWLRIPPIISPAQOI--Q	390						
Db	316	CEPEKFDVQVRNNSHLARGFTGHFCLGNOGLARLELSTLWTERLRLPLDLRLVADDSVPL	375						
QY	391	HKSGIVSGYQALPLYWDPRA	409						
Db	376	RPANFVSGIESMPVVTETPS	394						

```

RESULT      7
T30231
cytochrome P450 - Streptomyces hygroscopicus
C:Species: Streptomyces hygroscopicus
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 31-Jan-2000
C:Accession: J30231
R:Aparicio, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Straub
Gene 169, 9-16, 1996
A>Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hygroscopicus
A:Reference number: 220782
A:Accession: J30231
A>Status: preliminary; translated from GB/EMBL/DDBU
A:Molecule type: DNA
A:Residues: 1-404 <AAP>
A:Cross-references: EMBL:X86780; NID:g987088; PID:g987105; PIDN:CMA60465.1
C:Genetics:
A:Gene: rapM
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology

Query Match          15.9%; Score 346.5; DB 2; Length 404;
Best Local Similarity 28.6%; Pred. No. 5,76-19;
Matches 102; Conservative 63; Mismatches 177; Indels 15; Gaps 9;

OY    63 WATGGLREAYEDYRHFFSSECP---IPBAGAYDF-IPTSDPPEQGFALANOV 118
       | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db    50 WLASMEVAKFEDPR-FRSATLGKDPRLPAIQOPYMIMDDPEHTRLRRVATKA 108

OY    119 VGMPPVKLENNIOELACSLTESLRPOG-QCNFTDVAEPPIRIEFLMLAGLPEDIDPL 177
       | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db    109 LTRRMELARRTOGVADDLIDKMLAKGAPADMDFALPLITMICELLGYPIEDOTKF 168

OY    178 KYLDQMTRPGSKMT--FAEKREALDYLIPIEQRRQKPCTDAISIVANGQVGRDIT 234
       | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db    169 RTWSOML-SNGAVSQEVMAAGOSLYULSELIAERRKKOPTNDILGSLVARDKDDRL 227

OY    235 SDEAKRMCGLLVGGLDVVNPLFSWFEKLSPENHQELIERPERIPAAGEELLRRSL 294
       | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db    228 ETELGVFAVTLLIGSTETANAIGSYVTLTLHPKLAELKDSLIRKAVDELRIIPTI 287

OY    295 VADG---RLITSVDYEFHGVLCKGDQILLPMQLSGLDIERENACPMMHVDSROKVSHTFEG 351
       | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :

```

Dd	288 AKQASWRMAVEDVELSGTIVKAGEAVALQTHSANDDPKYVDHPEEIDPRTSNPHMSG	347
Qy	352 HGSILCLGQHLLARREIYTLTKEWLTRIPDSIA- PGAQLDHKSG- IYSGYQALPLWV	406
Dd	348 HGAIHCMAQVLRYEMQTALSGSLISRLPALRFAPVEPRIRFELGRULPSLSIALPLTW	404

RESULT 8  
B35401  
cytochrome P450 105B1 - Streptomyces griseolus  
N:contains: oxidoreductase (EC 1.-.-.-)  
C:Species: Streptomyces griseolus  
C:Date: 31-Mar-1992 #sequence.revision 31-Mar-1992 #text.change 04-Mar-2000  
C:Accession: B35401  
R:Omari, C.A.; Lestura, R.; Little, P.J.; Dean, C.; Tepperman, J.M.; Leto, K.J.; Romess  
J. Bacteriol. 172, 3335-3345, 1990  
A:Title: Genes for two herbicide-inducible cytochromes P-450 from Streptomyces griseo  
A:Reference number: A35401; MUID:90264332  
A:Accession: B35401  
A:Molecule type: DNA  
A:Residues: 1-403 <OMEGA>  
A:Cross-references: GB:M32239; NID:q153480; PIDN:AAA26825.1; PID:q153481; GB:M36481  
C:Genetics:  
A:Gene: CYP105B1  
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology  
C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygenase  
C:238-374/boman: cytochrome P450 homology <P45>  
C:352/Binding site: heme iron (Cys) (axial ligand) #status predicted

```

Query Match      15.6%  Score 340;  DB 2;  Length 403;
Best Local Similarity 28.9%;  Pred. No. 1,8e-18;
Matches 103;  Conservative 63;  Mismatches 165;  Indels 30;  Gaps 13;

QY      63  WATPFGQILREAYEDYRHFSSEC-----PFIPREAGEAYDFIP*-SMDPEQGRPALA 115
      111  : 111 : 111 : 111 : 111 : 111 : 111 : 111 : 111 : 111 :
Db      52  WLVTFRQDVRAVLDDPR-FSADAHRTGPFLLTAGRGELIGNPFLRMDDPEHARLRML 110
QY      116  NOVGMFPVYDKLENRIQELACSLIESLRP-OGGCFTEDEVAPPPRIIFMLAGLPEDI 174
      111  : 111 : 111 : 111 : 111 : 111 : 111 : 111 : 111 : 111 :
Db      111  TADFLVKKVEAMREPVQALADLDVLRMTGRTSADLVTEFALPLPSIVTCLLLGVPYEDH 170
QY      175  PHL-----KYLDDQMRPRGSGMTFAKAKALVDYLIPIIEGRKRGTDATS-IYANGVN 229
Db      171  AFPEGRSVALVLTLSKTP---EVRAQODELEYLARLRARKREPRDDAIIISRLVARELD 227
QY      230  GRPIITSDAEKRMCGLLLVGLDFTVYNELFSFMEFLANSPEHROELIERPERIPAACEEL 289
Db      228  DTQIAT-----MGRLLVAGHETTANMTALSTVLVLRNPQOLARLRAPALVKAVBEL 282
QY      290  RRFSLVADG--RIITSDYEFHGVOLKKGDQILLPOMLSGDERENACP--MHVDFSQKV 345
      111  : 111 : 111 : 111 : 111 : 111 : 111 : 111 : 111 : 111 :
Db      283  RYLTIIVHNGVPRIATEDVLIGRTIAGEGYLC--MISSANRDAEVPFGGDDLDVADAR 340
QY      346  SHTTFGHSHSLCLOHARREIIVTLKEMLRIRDPFSA-PGAQIOHKSGL-VSGVALP 403
      111  : 111 : 111 : 111 : 111 : 111 : 111 : 111 : 111 : 111 :
Db      341  RHVAFGEVHOCQPLARVELQAITETLLRLRDLRLAVPHEIIPFGMDATGVHSLP 400
QY      404  LVW 406
      : 1
Db      401  IAW 403

RESULT      9
O4BS6M
cytochrome P450 106 - Bacillus megaterium
N:Alternate names: cytochrome P450BM-1
N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Bacillus megaterium
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 03-Mar-2000
C:Accession: S07764; SI/973
R:He, J.S.; Ruettinger, R.T.; Liu, H.M.; Fulco, A.J.
J. Biochim. Biophys. Acta 1009, 301-303, 1989

```

A:Title: Molecular cloning, coding nucleotides and the deduced amino acid sequence of P  
A:Reference number: S07764; MUID:90089408  
A:Accession: S07764  
A:Molecule type: DNA  
A:Residues: 1-410 <HEJ1>  
A:Cross-references: EMBL:X1610; NID:g39626; PIDN:CA34612.1; PTD:g39627  
A:Accession: S17973  
A:Molecule type: protein  
A:Residues: 1-25 <HEJ2>  
C:Genetics:  
A:Gene: CYP106  
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology  
C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygenase;  
P:1-410/Product: cytochrome P450 106 #status experimental <MAT>  
F:241-378/Domain: cytochrome P450 homology <CYP>  
F:356/Binding site: heme iron (Cys) (axial ligand) #status predicted  
F:356/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match	15.3%;	Score 334.5;	DB 1;	Length 410;
Best Local Similarity	27.7%;	Pred. No. 4.8e-18;		
Matches	96;	Conservative 57;	Mismatches 170;	Indels 23;
				Gaps 7;

Oy 63 WIAIRGQILREAYEDYRHFSSE--CPETPRAGEAYDIP---TSMDEPQQRFRAIA 115  
| : : :: : | : : : | : : : : | : : : :  
Db 45 MNVEQYEHWKQLSNIDFSSSQGRRTITFGVDNSKKSKSPITNLTLNLDPPDRKRARSLL 104

```

QY      116  NQVGMGVVDKLENRQIOLACSLIESLRPQGCNFTEDYAEFPFRIIFMLAGLPEDIP 175
          :  ||:::  |:::  |  :  |::  |  :  |::  |
Db      105  AAAPPRSLKNWEPRIKQIADLVETAIKNSTINIVDDLSSPFPSLVIADLEGVPVKDRY 164

```

```

QY 176 HKLYTDQMPRPDGSMTFAEAE-----ALDYLLPIIEQRKKPGTAISIVANGQV 228
      | : : : | : : | : : : | : : : |
Db 165 QFKKWDILQPYDQERLEETIEQEKQAGAEYFQYLLPIVIEKRSNLSDDIISDLQAEV 224

```

**QY** 229 NGRPITSDEAKRMCGLLVGGDITVANFL-SFSMEFLAKSPENHRELIERERIPAAEE 287  
:  
:  
:  
**Db** 225 DGEFTTDEELIVATMILLGAGVETTSHTAIANMFYSFLYDDKSLYSELRNRELAPRAVEE 284

QY 288 LLR-RESLVADGRILTSDEYFNGVQLKKGGDQILLPQMSGLDERENACPMHIVDSFR-QKV 345  
::: ||: | : | | | | | | : : | | | |  
Db 285 MLRYRFHISKRDRTVAKQDNELGKVKLKKGGDVVYIAMSACNMDETMEFNPFSVDIHRPTNK 344

QY 346 SHTTFGHSJCLGQHARREIIVTLKNEWLTR-----IPDFSIA P 385  
| | | : | | | | | : : : : | | : |  
Db 345 KHLTFNGSPHFHCLGARPLARLEMKITLFAFLFARSHIEPEDFELEP 390

RESULT 10  
F70729  
cyclochrome P450 RV2266 - Mycobacterium tuberculosis (strain H37Rv)

N:contains:oxido-reductase (EC 1.-.-.-)  
C:Species: Mycobacterium tuberculosis  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999  
C:Accession: F70729  
C:Date: 10-Sep-1999 #text\_change 10-Sep-1999

McCoy, S.I.; Brosch, K.; Falkhuil, J.; Gathier, I.; Guichner, C.; Harris, J.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Rafanador, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares Nature 393, 537-544, 1998

A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the co  
A:Reference number: A70500; MUID:98295987  
A:Accession: F70729

A: Molecule type: DNA  
A: Residues: 1-428 <COL>  
A: Cross-references: GB:277163; GB:ALI23456; NID:g3261610; PIDN:CAB00969.1;

C:Genetics:  
A:Gene: RV2266  
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology

F;264-401/Domain: cytochrome P450 homology <CYP>

Query Match	15.3%;	Score 334.5;	DB 1;	Length 428;
Best Local Similarity	26.0%;	Pred. No. 5.1e-18;		
Matches 116;	Conservative 73;	Mismatches 188;	Indels 69;	Gaps 17;

```

5  IGSNANLAPLPVVEPHELVDFDMYNPSNLGAGVOEAMAVIQE-----SNVPDLY 54
   | : | | | | | : | : | : | : | : | : | : | : | : | : | : |
7  IATRVNGRP-PEEVP---IADIEL-----GSLDFMALDDDDVVRGAFATLRREAPISF 54

```

55 WTR-----CNGGHWMTRGOLIREAYEDY----RH---FSS-----ECPFIREA 92  
| ||| | : | ||| : | :  
55 WPTIELPGFVAGNGHWALTK-----YDDVFYASRHPDIFSSYPNITINDQTPPELAIEYF 107

93 GEAAYDFIPTSMDPPEQRFALANQVGMPIVDKLENRIGELACSLLESL---RPGQCN 149  
 108 GSM-----IVLDDPRHQRLRSIVSRAFTTPKVARIEAAVRRDRAHRLVSSMTANPNPQAD 162

```

150 FTEDYAEFPRIKFM LAGLPEDIPHLKYLTQNT--RDGSMFAEAKREALYD---Y 203
      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
163 LVSELAGPLPQIICDMMGIPKADHQRIEHWNTNVLIGFGDPDLATFDDEFMQVSADIGAY 222

```

204 LPIIEGRKRPETDAISIVANGVNGRPTSDKAIRMGCDLLVGDTVAVFSLFSNEEF 263  
 : | | : : | : | : | : | : | :  
 223 ATAAEDRVNHHDDLTSSLVEAEVDGERLSREIASFFILLVAGNETTRNAITHGVLA 282

204 LAKSPERQUELLERPERK - PAACEELLRRRS - LVAAGRILTSIDYEHGVQLKGDQLLLP 321  
::: || : | | ||:: | : | | | | : : ||:: |  
283 LSRYPEQRDRWNSDFDGLAFTAVEIIVRNASPVVYMRRTLQDIELRGTKMAAGDKSLW 342

343 YCSANRDESKFADPWTFLDARNPNPHLGFCGGCAIFCLGANLRDELIRVAEDELRRQMPD 402

361 FSIATGAGVYNNSGIVSGVALPLVM 400  
:  
: : : :  
403 V-VATEEPARLLSQFIHGIKTLPTW 427

RESULT 11  
69611  
cytochrome p450 cypA - Bacillus subtilis

```

;Accession: E69611; T44774
;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Feb-2000
;Species: Bacillus subtilis
;Accession: E69611; T44774

```

Bron, S.; Brouillette, S.; Bruschi, C.V.; Caldwell, B.; Capiano, V.; Carter, A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrante 390, 249-256, 1997

Leech, J.; Harwood, C.R.; Hernault, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hujoetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; L'Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.

Rieger, M.; Rivotta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.;  
Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska,  
Kuchni, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.;  
Victor, D.

Accession: E69611  
;Title: The complete genome sequence of the Gram-positive bacterium *Bacillus* s  
;Reference number: A69580; MUID:98044033

Cross-references: GR:299117, GR:AI009126, NID:02634966, PTDN:CA014615 1, PTD:  
;Residues: 1-410 <KUN>  
;Molecule type: DNA  
;Status: nucleic acid sequence not shown; translation not shown

**Title:** An *irpA*-like gene of *Bacillus subtilis* involved in branched-chain amino acid metabolism.

```

;Reference number: 4263/
;Accession: T44774
;Status: preliminary; translated from GB/EMBL/DBJ

```

A: Molecule type: DNA  
A: Residues: 1-410 <BEL>  
A: Cross-references: EMBL:Y110A3; PDB:1A97.1  
A: Experimental source: strain 1A1  
C: Genetics:  
A: Gene: cypA  
C: Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology  
C: Keywords: oxidoreductase  
C: 245-381/Domain: cytochrome P450 homology <CYP>

Query Match	15.38;	Score 333;	DB 1;	Length 410;
Best Local Similarity	27.28;	Pred. No. 6.2e-18;		
Matches	86;	Conservative 69;	Mismatches 139;	Indels 22;
			Gaps	5

[illegible]

```

Query Match          15.2%  Score 331.5:  DB 1:  Length 405:
Best Local Similarity 25.6%:  Pred. No. 8e-18;
Matches 107:  Conservative 73:  Mismatches 185;  Indels 53:  Gaps 10.

QY 18 VPEHLVDP-----FDMYNSNTLSAGVOEAMAYLQESNVPDLVWTRCNG-CHWIAIRKQLIR 72
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
6 VPDLLAFGEADFAQDPHN-----RYARRKEEPVORI-RTVNGDGLDAWITLRDEYK 53
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

```

QY 73 EAYEDYHHFSESCECFIPREAGEAVIDPISPM-----DPEQROFR 112
Db 54 QALLD-----PRIAKDGRQOIIIEKRLADERRRPFSDLCRPHMLNMDPDHFLR 105
QY 113 ALANQVGMPIVDEKLEIRIDELACLSIESLRPGOGCNFTEDYAEBPPIRIFMLLAGDEE 172
Db 106 KLVYKATTAARVEELRRRIQIITDDLLDRLAGRSEVDLIDERAFPLPTIVISELGAVDS 165
QY 173 DIPHKLKLTDOMTPRDGSMTFAAKE--ALYUDLPYIEOROKGPTAISIVANGQNG 230
Db 166 RRDFRSMWTNVLV--DSSGEQAQAQASVAMVELTLELAKKTBEGDDLLTALLEAVDGS 223
QY 231 RPTISDEAKRRCGLLVGLDITYVNFLSFSMEFLAKSPHROELIERPERIPACCELLR 290
Db 224 DRLSEGETLIMVFLLVAGHETTVNLIGNCVSLIGNPQJLAALRNDSLLGALIEETLR 283
QY 291 RFLSLVADG--RILTSYEFHGVOLKKGDDQILLPQMLSGLDERENAKCPMHVDSROKYSHT 348
Db 284 YESYVANGTFRHTAAEAVERFQDVVYIPGSELVMAVGANRDXGERFEDPRDFITRETTGHW 343
QY 349 TEFGSHLCLGONHILAREELIVLKEMLITRIPDESTA--GCAITQKSGISVSYQALPL 404
Db 344 AFGGHHFYCGAALRLAEQIIVAGRLLERPPDLRMAASDDLRWRFVSLMRLEKLPV 401

```

Query Match	15.18;	Score 330;	DB 2;	Length 406;
Best Local Similarity	26.38;	Pred. No. 1e-17;		
Matches 104;	Conservative 74;	Mismatches 193;	Indels 24;	Gaps 11;
QY	31	PSNLSAGVQ--EAMVAIDESNVPLDVTWRCNGH-WIATRCQLIREAVEDYR-----	79	
Db	17	PSNNSCPQLQEDGVAKQLDTPGRLHRYTLTYGROAMVTKHEAARKLLGDRLLSNRTDD	76	
QY	80	HFSSECEPIP--REAGEAYDFIPTSMDDPEOROFALANQVYVGMVVDKLENRIOALACS	137	
Db	77	NFPATSPFEAVRESPQAF-----IGLDPPENGTRRMNITSEFTYKRIKGMPVEVEVYHG	132	
QY	138	LIESLRPGQ-OCNFTEDYAEPPRIPIEMLLAGLPEDIPHIKLYLTDDQTRPGSMTEAEA	196	
Db	133	FLDEMLAGPADIADLVSGQALPVPMSVICRLGLGYVAHDFEPQDASKRLVQSTDAQSALTA	192	
QY	197	KEALVDYILPIIEQRQKPGTDAT-SIVANGVNGRPTSDSEAKRMCGLLIVGGLDTVVN	255	
Db	193	RNDLAGYLDGLITQFQTEPGAGALVGAIVADQLANGE-IDREELISTAMLLIAGHETTAS	251	
QY	256	FLSPSMETLASPEHROILIRPERIPACCELLRRSL--VADGRILTSYEPHQVLK	313	
Db	252	MTSLSVITLLDHPDYALRADRSILPVGAVETELLRYLAIADISAGRAVATDIEVEGHILIR	311	





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 12:36:49 ; Search time 27.02 Seconds  
(Without alignments)  
474.815 Million cell updates/sec

Title: US-09-246-451-2

Perfect score: 2180  
Sequence: 1 TTETIQSNANLAPLPHPVPE.....IVSGVOALPLVWDPATTKAV 414

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 85661 seqs, 30989116 residues

Total number of hits satisfying chosen parameters: 85661

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_38.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2180	100.0	414	1 CPXA_PSEPU	P00183 pseudomonas
2	595.5	27.3	414	1 Y4VG_RHISN	053215 rhizobium s
3	380	17.4	393	1 YH85_MYCTU	053926 mycobacteri
4	371	17.0	410	1 CPXM_BACME	006069 bacillus me
5	362	16.6	396	1 YJTB_BACSU	034374 bacillus su
6	360	16.5	398	1 YZ18_MYCTU	053563 mycobacteri
7	357.5	16.4	405	1 CPXM_BACSU	P27632 bacillus su
8	340	15.6	402	1 CPXF_STRGO	P18327 streptomyc
9	334.5	15.3	410	1 CPXF_BACME	P14762 bacillus me
10	334.5	15.3	428	1 YM66_MYCTU	050696 mycobacteri
11	333	15.3	410	1 CPXG_BACSU	008469 bacillus su
12	331.5	15.2	405	1 CPXK_SACER	P33271 saccharopol
13	330	15.1	405	1 CPXK_STRGO	P18327 streptomyc
14	320.5	14.7	405	1 YC56_MYCTU	011062 mycobacteri
15	318.5	14.6	395	1 BIOT_BACSU	P53554 bacillus su
16	318	14.6	428	1 CPXL_PSEPU	P33006 pseudomonas
17	316.5	14.5	399	1 FASI_RHOFA	P46373 rhodococcus
18	316	14.5	412	1 CPXK_STRGO	P26911 streptomyc
19	304	13.9	438	1 Y180_MYCTU	008464 mycobacteri
20	303	13.9	402	1 NOR_FUSOX	P23295 fusarium ox
21	296.5	13.6	408	1 NOR_CYLTO	012599 cyllindroc
22	293	13.4	401	1 CPXG_STRGO	P23296 streptomyc
23	291.5	13.4	433	1 YZ45_MYCTU	P71856 mycobacteri
24	289.5	13.3	404	1 CPXJ_SACER	000441 saccharopol
25	288.5	13.2	403	1 NOR_CYLTO	000464 cyllindroc
26	277.5	12.7	397	1 CPXO_SACER	P48635 saccharopol
27	266	12.2	400	1 CPXP_RHISN	P55544 rhizobium s
28	265	12.2	414	1 Y778_MYCTU	P77993 bradyrhizob
29	254	11.7	401	1 CPXP_BRAJA	059203 mycobacteri
30	236	10.8	402	1 YM76_MYCTU	P77992 mycobacteri
31	235	10.8	396	1 YM76_MYCTU	059571 mycobacteri
32	224	10.3	436	1 THCB_RHOER	P43492 rhodococcus
33	221.5	10.2	400	1 YV21_MYCTU	008362 mycobacteri

34	218	10.0	422	1 CPXC_AGR6	P24466 agrobacteri
35	215	9.9	429	1 CPXR_BRAJA	059204 bradyrhizob
36	207.5	9.5	489	1 YMP8_MYCTU	059572 mycobacteri
37	195	8.9	405	1 CPXP_BACSU	034926 bacillus su
38	189	8.7	467	1 CPXK_RHISN	P55543 rhizobium s
39	181	8.3	368	1 CPXM_SULISO	055080 sulfolobus
40	162	7.4	1048	1 CPXB_BACME	P14779 bacillus me
41	156.5	7.2	503	1 CP30_SHEEP	029496 ovis aries
42	153.5	7.0	531	1 CP27_HUMAN	002318 h cytochrom
43	152	7.0	492	1 CP53_PIG	002390 sus scrofa
44	151	6.9	407	1 CPXD_AGR6	P24467 agrobacteri
45	149.5	6.9	489	1 CP53_RABIT	P00182 oryctolagus

## ALIGNMENTS

RESULT	1	STANDARD	PRT	414 AA.
CPXA_PSEPU				
ID	CPXA_PSEPU			
AC	P00183			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	13-AUG-1987 (Rel. 05, Last sequence update)			
DT	15-FEB-2000 (Rel. 39, Last annotation update)			
DE	CYTCHROME P450-CAM (EC 1.14.15.1) (CAMPHOR 5-MONOOXYGENASE).			
GN	CAMC OR CYP101.			
OS	Pseudomonas putida.			
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;			
OC	Pseudomonas.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE: 86111751.			
RA	Unger B.P., Gunsalus I.C., Slinger S.G.:			
RT	"Nucleotide sequence of the Pseudomonas putida cytochrome P-450cam			
RT	gene and its expression in Escherichia coli."			
RL	J. Biol. Chem. 261:1158-1163(1986).			
RN	[2]			
RP	SEQUENCE OF 385-414 FROM N.A.			
RC	SPRAIN-ATCC 17453:			
RX	MEDLINE: 90130389.			
RA	Koga H., Yamaguchi E., Matsunaga K., Aramaki H., Horiuchi T.:			
RT	"Cloning and nucleotide sequences of NADH-putidaredoxin reductase			
RT	gene (camA) and putidaredoxin gene (camB) involved in cytochrome			
RT	P-450cam hydroxylase of Pseudomonas putida."			
RL	J. Biochem. 106:831-836(1989).			
RN	[3]			
RP	SEQUENCE:			
RX	MEDLINE: 83030788.			
RA	Hanlin M., Ames L.G., Yasunobu K.T., Shastry B.A., Gunsalus I.C.:			
RT	"Amino acid sequence of the Pseudomonas putida cytochrome P-450. II.			
RT	Cyanogen bromide peptides, acid cleavage peptides, and the complete			
RT	sequence."			
RL	J. Biol. Chem. 257:12664-12671(1982).			
RN	[4]			
RP	X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).			
RX	MEDLINE: 86143817.			
RA	Fitzel B.C., Weber P.C., Hardman K.D., Saleme F.R.:			
RT	"Structure of ferricytochrome c' from Rhodospirillum molischianum at			
RT	1.67-A resolution."			
RL	J. Mol. Biol. 186:627-643(1985).			
RN	[5]			
RP	X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).			
RX	MEDLINE: 86059514.			
RA	Poulos T.L., Fitzel B.C., Gunsalus I.C., Wagner G.C., Kraut J.:			
RT	"The 2.6-A crystal structure of Pseudomonas putida cytochrome P-450."			
RL	J. Biol. Chem. 260:16122-16130(1985).			
RN	[6]			
RP	X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).			
RX	MEDLINE: 98019009.			
RA	Schlichting I., Jung C., Schulze H.:			
RT	"Crystal structure of cytochrome P-450cam complexed with the (1S)-			
RT	camphor enantiomer."			
RL	FEBS Lett. 415:253-257(1997).			

[7]  
RN X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).  
RP MEDLINE: 98313255.  
RA Vidakovic M., Slijar S.G., Li H., Poulos T.L.:  
RT "Understanding the role of the essential Asp51 in cytochrome p450cam  
RT using site-directed mutagenesis, crystallography, and kinetic solvent  
RT isotope effect.";  
RL Biochemistry 37:9211-9219(1998).  
RN  
RP STRUCTURE BY NMR.  
RX MEDLINE: 97459726.  
RA Moura C., Bondon A., Simmoneaux G., Jung C.:  
RT "H-NMR study of diamagnetic cytochrome P450cam: assignment of heme  
RT resonances and substrate dependence of one cysteinate beta proton.";  
RL FEBS Lett. 414:203-208(1997).  
CC  
CC -1- FUNCTION: INVOLVED IN A CAMPHOR OXIDATION SYSTEM.  
CC -1- CATALYTIC ACTIVITY: CAMPHOR + PUTIDAREDOXIN + O(2) = 5-EXO-  
CC HYDROXYCAMPHOR + OXIDIZED PUTIDAREDOXIN + H(2)O.  
CC  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: M12546; AAA25760.1; -;  
DR EMBL: D00528; BAA00412.1; -;  
DR PIR: A25660; QAPSCP.  
DR PDB: 2COP; 15-APR-91.  
DR PDB: 3COP; 15-APR-91.  
DR PDB: 4COP; 15-JUL-91.  
DR PDB: 5COP; 15-JUL-91.  
DR PDB: 6COP; 15-JUL-91.  
DR PDB: 7COP; 15-JUL-91.  
DR PDB: 8COP; 15-JUL-91.  
DR PDB: 1CP4; 15-JAN-93.  
DR PDB: 2CP4; 15-JAN-93.  
DR PDB: 3CP4; 15-JAN-93.  
DR PDB: 4CP4; 15-JAN-93.  
DR PDB: 5CP4; 16-SEP-98.  
DR PDB: 6CP4; 16-SEP-98.  
DR PDB: 1NOO; 08-MAR-96.  
DR PDB: 1PHA; 31-OCT-93.  
DR PDB: 1PHB; 31-OCT-93.  
DR PDB: 1PHC; 31-OCT-93.  
DR PDB: 1PHD; 31-OCT-93.  
DR PDB: 1PHE; 31-OCT-93.  
DR PDB: 1PHE; 31-OCT-93.  
DR PDB: 1PHG; 31-OCT-93.  
DR PDB: 1ARD; 19-NOV-97.  
DR PRAM: PF00067; p450; 1.  
DR PROSITE: PS00086; CYTOCHROME\_P450; 1.  
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;  
KW 3D-structure.  
FT INIT MET 0  
FT BINDING 357 357 HEME.  
FT CONFLICT 55 56 MISSING (IN REF. 3).  
FT CONFLICT 276 276 E -> Q (IN REF. 3).  
FT CONFLICT 361 361 H -> S (IN REF. 3).  
FT CONFLICT 407 407 D -> N (IN REF. 3).  
FT TURN 16 17  
FT HELIX 20 22  
FT STRAND 23 23  
FT TURN 28 29  
FT TURN 32 33  
FT HELIX 34 36  
FT HELIX 38 46  
FT TURN 48 49  
FT STRAND 53 56  
FT HELIX 58 60

FT STRAND 62 65  
FT HELIX 68 76  
FT TURN 78 80  
FT STRAND 81 82  
FT TURN 83 84  
FT HELIX 90 95  
FT TURN 99 102  
FT TURN 105 107  
FT HELIX 108 119  
FT HELIX 121 145  
FT TURN 146 146  
FT STRAND 147 149  
FT HELIX 150 153  
FT TURN 154 156  
FT HELIX 167 167  
FT TURN 168 168  
FT HELIX 171 185  
FT HELIX 193 213  
FT HELIX 219 224  
FT TURN 225 225  
FT STRAND 227 228  
FT TURN 229 230  
FT STRAND 231 232  
FT HELIX 235 250  
FT TURN 251 252  
FT HELIX 253 266  
FT HELIX 268 276  
FT HELIX 278 291  
FT TURN 292 292  
FT STRAND 295 295  
FT STRAND 297 301  
FT TURN 305 307  
FT TURN 308 309  
FT STRAND 310 312  
FT TURN 314 315  
FT STRAND 317 320  
FT HELIX 322 324  
FT TURN 325 327  
FT TURN 329 331  
FT TURN 335 336  
FT TURN 340 341  
FT TURN 349 350  
FT HELIX 353 355  
FT TURN 358 359  
FT HELIX 360 377  
FT STRAND 382 383  
FT TURN 385 386  
FT STRAND 391 392  
FT STRAND 396 396  
FT STRAND 398 399  
FT STRAND 403 405  
FT HELIX 408 410  
SQ SEQUENCE 414 AA; 46538 MW; ECA610293A9D6207 CRC64;  
  
Query Match 100.0%; Score 2180; DB 1; Length 414;  
Best local Similarity 100.0%; Pred. No. 7.4e-159;  
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 TTETIOSNANLAPLPHVPHVLFVDFDQMYNPSNLSAGVQEAAMVLDSENVPDLVWTRCNG 60  
DB 1 TTETIOSNANLAPLPHVPHVLFVDFDQMYNPSNLSAGVQEAAMVLDSENVPDLVWTRCNG 60  
QY 61 GHWIATRGOLIRFAYDYRHFSSECFPIPREAGEAYDFTPTSMDDPPEOROFALANQVVG 120  
DB 61 GHWIATRGOLIRFAYDYRHFSSECFPIPREAGEAYDFTPTSMDDPPEOROFALANQVVG 120  
QY 121 MPVVDKLENRIOELACLSESLRPQGCNFTEDYAEFPPIRIFMLAGLPEEDIPIHLKYL 180  
DB 121 MPVVDKLENRIOELACLSESLRPQGCNFTEDYAEFPPIRIFMLAGLPEEDIPIHLKYL 180  
QY 181 TDQMTRPDGSMTFAEAKKALYDLPIPIEOROKPGTDAISIVANQVNGRPITSDEAKR 240  
DB 181 TDQMTRPDGSMTFAEAKKALYDLPIPIEOROKPGTDAISIVANQVNGRPITSDEAKR 240



Db	181	TDOMRPRGSMF	FAAKALDYLLIP	ITEGRQKPGCTA	ISIVANGQVNGREITSDEAKR	240
Qy	241	MGCLLVGGDLTVN	WFSFSMEFLAKSP	EHQELLERPERITPA	ACEDELLRRSLVADGR	300
Db	241	MGCLLVGGDLTVN	WFSFSMEFLAKSP	EHQELLERPERITPA	ACEDELLRRSLVADGR	300
Qy	301	LTSDYEPFAGVOL	KKDDQILLPOMLSG	LDERENACPMHNDP	FSROKVSHTTFFGSHL	CGQ 360
Db	301	LTSDYEPFAGVOL	KKDDQILLPOMLSG	LDERENACPMHNDP	FSROKVSHTTFFGSHL	CGQ 360
Qy	361	HLAREIIVTLKEW	LTRIPDFSIA	GAQIQHKSIVSGV	QVALPLVMDPATTKAV	414
Db	361	HLAREIIVTLKEW	LTRIPDFSIA	GAQIQHKSIVSGV	QVALPLVMDPATTKAV	414
RESULT	2					
ID	Y4VG_RHISN	STANDARD:	PRT:	414	AA.	
AC	Y4VG_RHISN					
DT	01-NOV-1997	(Rel. 35, Created)				
DT	01-NOV-1997	(Rel. 35, Last sequence update)				
DT	13-DEC-1998	(Rel. 37, Last annotation update)				
DE	PROBABLE CYTOCHROME P450 Y4VG (EC 1.14.14.-).					
GN	Y4VG.					
OS	Rhizobium sp. (strain NGR234).					
OC	Plasmid sym PNGR234a.					
OC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;					
NC	Rhizobiaceae; Rhizodium.					
RA	SEQUENCE FROM N.A.					
RA	Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,					
RA	Perret X.:					
RT	"Molecular basis of symbiosis between Rhizodium and legumes.";					
RL	Nature 387:394-401(1997).					
RN	[2]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE: 96389014.					
RA	Freiberg C., Perret X., Broughton W.J., Rosenthal A.:					
RT	"Sequencing the 500-kb GC-rich symbiotic replicon of Rhizodium sp.					
RT	NGR234 using dye terminators and a thermostable 'sequenase': a					
RT	beginning.";					
RL	Genome Res. 6:590-600(1996).					
CC	-1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE					
CC	MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED					
CC	COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.					
CC	-1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.					
CC	-----					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration					
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -					
CC	the European Bioinformatics Institute. There are no restrictions on its					
CC	use by non-profit institutions as long as its content is in no way					
CC	modified and this statement is not removed. Usage by and for commercial					
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> ).					
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).					
CC	-----					
DR	EMBL: Z68203: CAA92422.1: -					
DR	EMBL: AE000101: AAB91895.1: -					
DR	HSSP: P00183: 3CP4.					
DR	PFAM: PF00067: P450_2.					
DR	PROSITE: PS00086: CYTOCHROME_P450; FALSE NEG.					
KW	Hypothetical protein; Oxidoreductase; Monooxygenase;					
KW	Electron transport; Heme; Plasmid					
FT	BINDING 364 HEME (BY SIMILARITY).					
SE	SEQUENCE 414 AA: 45810 MW: 888598E99315BB84 CRC64:					

```

Query Match      27.3%   Score 595.5; DB 1; Length 414;
Best Local Similarity 34.38%; Pred. No. 3.4e-38;
Matches 143; Conservative 64; Mismatches 171; Indels 39; Gaps 9;

OY    13 PLPRVPEHLVDPMYNPSNLSA---GYQEWAVLVQESNVLDYW----TRCGNGHRIA 65
      ||| |||| | | : : : | : : : : | | | | :

```

```

Db      13 PIPDHVPALVRHFSLETPSPOMATTPNGDHAACVCHDDGCPPIFYSPSNTBDGRGHWI  72
Qy      66 TRGQIREAVEDYHFSSECFIPREAGEADYLPITSMDDPEORORFALANOVYGMPPVD  125
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      73 TRARDORVLEDETETFFSHRSIFASALGEHMPVPIPLDLPDPAHGVFALNPLPFSSARVY  132
Qy      126 KLERIDELACSLIESI-RPGGOCNPFEDYDAEPPRIIRIFMLAGLPR-----EDI  174
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      133 ALEPTIHARAGALIDLCJAKETSCDVKMDRALPPTFESVFLSGLSORSEVLGVWSDI  192
Qy      175 PHLRYLTDMORPMPSMTPEAKKALDYLIPIEEQ-----RROKPGDAISIAVANGG  230
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      193 LH-----GN---AEKRAARASVAFIDEMAARRRSPANDFTEFYVQAKIEG  237
Qy      231 RPTISDEAKRMCGLLLVGGLDVTYVNLFSFNEFLAKSPHEHQELIERPERIPACEELLR  290
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      238 RSLTEEEVRGIGVFLVAGGLDTVAALIGFDMAYLARNPKHQELLRNBPARTGLAAEELLR  297
Qy      291 RESLVADGRILTSYDEHGVQVKKGGDOLLPMQLSGIDEDENACPMHVPDSROKVSHTTE  350
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      298 AYSTVOIIRVATKDIIEEGVPIREGDVSVPAMLANDBPEEFCOPNTIDARODNHTAF  357
Qy      351 GGGSHLGLGOHLAREIIVTLKEVLTIPPESTAPG-NQLOHSGIYSGVQALPLWV  406
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      358 GYGHLCGHAHLARREIYIGIREMLAIPAPRIKEGAPLTH-GHGVFGISNIIITW  413

```

RESULT	3
YH85_MYCTU	
ID	YH85_MYCTU
AC	053936;
DT	15-FEB-2000 (Rel. 39, Created)
DT	15-FEB-2000 (Rel. 39, Last sequence update)
DT	15-FEB-2000 (Rel. 39, Last annotation update)
DE	PUTATIVE CYTOCHROME P450 RV1785C (EC 1.14.-.-).
GN	RV1785C OR MYV04.9.07C.
OS	Mycobacterium tuberculosis.
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN	(1)
RP	SEQUENCE FROM N.A.
RC	SPRAIN-H37RV.
RC	MEDLINE; 98295987.
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA	Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA	Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R., Stulton J.E.,
RA	Taylor K., Whitehead S., Barrall B.G.;
RT	"Deciphering the biology of Mycobacterium tuberculosis from the
RL	complete genome sequence."
RL	Nature 393:537-544(1998).
CC	-I- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL; AL022021; CAA17707.1; -.
DR	TUBERCULIST; RV1785C; -.
DR	PFAM; PF00067; P450; 2.
DR	PROSITE; PS00086; CYTOCHROME_P450; 1.
KW	Hypothetical protein: Oxidoreductase; Monooxygenase; Heme.
FT	BINDING 342 342 HEME (BY SIMILARITY).
SO	SEQUENCE 393 AA; 43541 MW; 8BCDD150B471FAB8 CRC64;

DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
 KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.  
 FT BINDING 342 342 HEME (BY SIMILARITY).  
 SO SEQUENCE 393 AA; 43541 MM; BCCDD5DB47FAB8 CRC64;

```

Query Match          17.4%; Score 380; DB 1; Length 393;
Best Local Similarity 28.7%; Pred. No. 8,1e-22;
Matches 110; Conservative 71; Mismatches 182; Indels 20; Gaps 10;

QY 31 PSLSLAGVQAMAVLQESNPDLVTRCNGHVIATRGQIREAYEDYRHFSSECPPIPR 90
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 23 PMAVDGV--GMKTLRDAG--PVVF--MNCWVYL--TRREVIYLAALRNPKEVSSRKALQ- 74
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 91 EAGEAVDFIPTSDPPEOROFALANOVGMVYDKLENRIQELACLSISLSPGOCNF 150
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 75 -PKNPLPVYPLADPPREHTRKRIILOPYSPALSKALPRLRHRYAMIDALINGRCEA 133
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 151 TEDYAPFPRIEPMILAGLPREDIPHLKYITDQ---TRPDGSMFAEAKALVDYLLP 206
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 134 MADLANLPFQLFLVYLGLPLEDRDLIGKWDVAMSDRPHPTADVAAARELLEYLTA 193
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 207 IIGORRQPGTDAISIVANQVGRPITSDEAKRMCGLLVGLDIPVYNLTSMEFLAK 266
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 194 MVAERRRNPQDVLQOV--QIGEDPLSEIEVLGSLHLITLADIDVYTAAGVFSLELAR 250
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 267 SPEHROELIERPERIPACBELLR-RFSVADGRILTSDEFFGVOLKKGDOILLPQMLS 325
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 251 RQPLRLMRNPQOIVFTEIYRLERSAPVAVRVTTEPTVGGMTLPAGSPVRLCMAY 310
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 326 GIDERENACPMHVDFSRQKVSHTTFGHSHLCLGQHLARREIIVTLKEMWLRIPDSIAP 385
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 311 NRDSQAMSTDELVMQKVRHMGFGGPHRCISHLARLETLVLGEMWLNQJPDEFELAP 370
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 386 --GAQIOHKSIGVYGVQALPLVW 406
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 371 DYAPETIRFSPKSPA-LKNLPLRW 392
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 4
CPXM_BACME          STANDARD;          PRT;          410 AA.
ID 006069;
AC 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE CYTOCHROME P450(MEG) (EC 1.14.99.-) (STEROID 15-BETA-HYDROXYLASE)
DE (STEROID 15-BETA-MONOOXYGENASE).
GN CYP106A2.
OS Bacillus megaterium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13368;
RX MEDLINE: 94049677.
RA Rauschenbach R., Isernhagen M., Noeske-Jungblut C., Boidol W.,
RA Stewert G.;
RT "Cloning sequencing and expression of the gene for cytochrome
RT P450meg, the steroid-15 beta-monooxygenase from Bacillus megaterium
RT ATCC 13368.";
RT Mol. Gen. Genet. 241:170-176(1993).
RN [2]
RP CHARACTERIZATION.
RC STRAIN=ATCC 13368;
RX MEDLINE: 79194051.
RA Berg A., Ingelman-Sundberg M., Gustafsson M.;
RT "Purification and characterization of cytochrome P-450meg.";
RT J. Biol. Chem. 254:5264-5271(1979).
RN [3]
RP CHARACTERIZATION.
RC STRAIN=ATCC 13368;
RX MEDLINE: 82091079.
RA Berg A., Rafter J.J.;
RT "Studies on the substrate specificity and inducibility of cytochrome
RT P-450meg.";
RT Biochem. J. 196:781-786(1981).
CC -1- FUNCTION: HAS THE CAPACITY TO HYDROXYLATE CERTAIN STEROIDS IN THE
CC 15-BETA POSITION.

```

```

CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: Z21972; CA91985.1; -.
DR PIR: S32216; S32216.
DR PIR: S39924; S39924.
DR PFAM: PF00067; p450; 1.
DR PRINTS: PR00359; BP450.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.
FT BINDING 355 HEME (BY SIMILARITY).
SQ SEQUENCE 410 AA; 46955 MW; 9FDD9CF2E80F810B CRC64;

Query Match          17.0%; Score 371; DB 1; Length 410;
Best Local Similarity 30.1%; Pred. No. 4,1e-21;
Matches 109; Conservative 53; Mismatches 180; Indels 20; Gaps 7;

QY 63 WIAITRQGLREAYEDYRHFSSE-----CPFTIRFAGEAVDFIP-TSMDPEQORFALA 115
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 44 MWVFYEDYKRVLSQDKHFTSSVAKRTTISVGTSEGSVPEKQIIESPPDRKRRL 103
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 116 NOVGMVYVDKLENRIQELACLSIESLRPOGOCNTEYDAEPPIRIFMILAGLPREDIP 175
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 104 AAATFPRSLQNMPEIRIQETIADELIGQMDGTEIDIVASLASPLPIIYMDLIMVPSKRL 163
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 176 HLKYITDQMTRPDGSMTPE-----AKEALDYLLPIIEQRORQPGTDAISIVANGQV 228
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 164 LFKKWVDPLFLPDRKQOEVDKLRQVAAKEYQYLYPIYQKRLNPADDIISDLKSEV 223
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 229 NGRPITSDEAKRMCGLLVGLDIPVNFLSFSM-EFLAKSPEHROELIERPERIPACSE 287
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 224 DGEMLFDDEVRTMILLAGVETTSHLANSFYSILYDKKEVYQELHENDLVPQAVEE 283
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 288 LIR-RFSVADGRILTSDEFFGVOLKKGQILLPQMLSGIDERENACPMHVDFSR-QKV 345
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 284 MLRFENFLIKIDRTVKEEDNDLGLVELKEGDSVYVMMASAMDEMEDEPFTLIHRPNK 343
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 346 SHRTFGHSHLCLGQHLARREIIVTLKEMWLRIPDSIAPGNOIQRK---SGIVSGVAL 402
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 344 KHLTFNGRPHFLGAPLARLEAKIALTAFLKFKKHTAAPSFOLENLTDSATGQTLTSL 403
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 403 PL 404
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 404 PL 405
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 5
YJIB_BACSU          STANDARD;          PRT;          396 AA.
ID YJIB_BACSU
AC 034374;
DT 15-FEB-2000 (Rel. 39, Created)
DT 15-FEB-2000 (Rel. 39, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE PUTATIVE CYTOCHROME P450 YJIB (EC 1.14.-.-).
GN YJIB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Rivolta C., Soldo B., Lazarevic V., Joris B., Manuel C., Karamata D.;
RT "A 35.7 kb DNA fragment from Bacillus subtilis chromosome containing a
RT putative 12.3 kb operon involved in hexuronate catabolism and a
RT perfect catabolite-responsive element.";

```





```
AC P14762;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE CYTOCHROME P450(BM-1) (EC 1.14.14.1).
GN CYP106.
OS Bacillus megaterium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-25.
RC STRAIN=ATCC 14581;
RX MEDLINE; 90089408.
RA He J.S., Ruettiger R.T., Liu H.-M., Fulco A.J.;
RT "Molecular cloning, coding nucleotides and the deduced amino acid
RT sequence of P-450BM-1 from Bacillus megaterium.";
RL Biochim. Biophys. Acta 1009:301-303(1989).
RN [2]
RP SEQUENCE OF 1-25 FROM N.A.
RX MEDLINE; 95355495.
RA He J.S., Liang O., Fulco A.J.;
RT "The molecular cloning and characterization of BM1p1 and BM1p2
RT proteins, putative positive transcription factors involved in
RT barbiturate-mediated induction of the genes encoding cytochrome
RT P450BM-1 of Bacillus megaterium.";
RL J. Biol. Chem. 270:18615-18625(1995).
CC -1 FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
CC MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED
CC COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.
CC -1 SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X16610; CAA34612.1; -
DR EMBL: S79230; AAC60495.1; -.
DR PIR: S07764; O4B56M.
DR HSSP: P33006; 1CPT.
DR PFAM: PF00067; P450; 1.
DR PRINTS: PR00359; BP450.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
DR Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.
KM BINDING 356 356 HEME.
FT SEQUENCE 410 AA; 47460 MW; C9AE293E76745387 CRC64;
SQ

Query Match 15.3%; Score 334.5; DB 1; Length 410;
Best Local Similarity 27.7%; Pred. No. 2.5e-18;
Matches 96; Conservative 57; Mismatches 170; Indels 23; Gaps 7;

63 WATRGQLREAVEDYHFSSE--CPFPRAGEAVDFIP---TSMDEPQKQFALA 115
+ : : : : : : : : : : : : : : : : : : : : : : : : : :
45 WNFQYHVVQVLSNYFFSSDGGRTTIFVGDNSKKKSTPTNLNLDPDRHKKARSL 104
+ : : : : : : : : : : : : : : : : : : : : : : : : : :
116 NOVVGMPVDKLENRIQELACSLIESLRPOGOCFTEDVAEPPIRIFMLAGLPEEDIP 175
+ : : : : : : : : : : : : : : : : : : : : : : : : : :
105 AAATPRSLKNWEPRIKQIADLVEAIQKNSTINIVDLSPPFSLVADLFQVPAVDY 164
+ : : : : : : : : : : : : : : : : : : : : : : : : : :
176 HLKYLDDQTRPQGSMTFAAKE-----ALDYLIPIIEQRQKGTALISVANGQV 228
+ : : : : : : : : : : : : : : : : : : : : : : : : : :
165 QFKRWVDILFQVDERLEIEQEKORAGAEYQYLLPIVIEKSNLSDDIISDLQAEV 224
+ : : : : : : : : : : : : : : : : : : : : : : : : : :
229 NGRPITSDEAKRMGGLLVGLDPTVNFLLSFSMEFLAKPPEHROELIERPERIPACEE 287
+ : : : : : : : : : : : : : : : : : : : : : : : : : :
225 DGEFTDEELVHATMLLGGVETTSALINMFTSYLYDKSKLSYSELRNRELAPKAVEE 284
+ : : : : : : : : : : : : : : : : : : : : : : : : : :
288 LLR-RFSLVADGRLTSDYEFHGVQLKKQDQILLPQMLSGIDEREKACPMHVDFSR-QKV 345
+ : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
DB 285 MLRYRPHISRRDRTVKODNELLGVKLRKGDVYIAMSACNMDETMEPNPSPVDIHRPTNK 344
OY 346 SHPTFGHSHLCIGQHLAREIIVTLKEWLR-----IPDEFIAP 385
+ : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 345 KHLTFGNPHFCUGAPLARLEMKITIEAFLEAFSHIEPFEFDELEP 390
+ : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
YME6_MYCTU
ID YME6_MYCTU STANDARD: PRT; 428 AA.
AC Q50696;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 15-FEB-2000 (Rel. 39, Last annotation update)
DE PUTATIVE CYTOCHROME P450 RV2266 (EC 1.14.-.-).
GN RV2266 OR MYCY339.44C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE; 98295987.
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule J., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
CC -1 SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z77163; CAB00969.1; -.
DR HSSP: P33006; 1CPT.
DR TUBERCULIST: RV2266; -.
DR PFAM: PF00067; P450; 1.
DR PRINTS: PR00359; BP450.
DR PROSITE: PS00086; CYTOCHROME_P450; FALSE_NEG.
DR Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.
KM BINDING 379 379 HEME (BY SIMILARITY).
FT SEQUENCE 428 AA; 47824 MW; 76B1FC5AE348591 CRC64;
SQ

Query Match 15.3%; Score 334.5; DB 1; Length 428;
Best Local Similarity 26.0%; Pred. No. 2.6e-18;
Matches 116; Conservative 73; Mismatches 188; Indels 69; Gaps 17;

5 IQSNANLAPRPPIPHLAVDFDPMYNPSNLSAGVQEAVALQOE-----SNVDPDLV 54
+ : : : : : : : : : : : : : : : : : : : : : : : : : :
7 IATRVNQT-P-PEVP---IADIEL-----GSLDWALDDVDROCAFATLRREAPISF 54
+ : : : : : : : : : : : : : : : : : : : : : : : : : :
55 WTR-----CNGGHWTATRGQLIREAYEDY---RH---FSS-----ECQFIPREA 92
+ : : : : : : : : : : : : : : : : : : : : : : : : : :
55 WPTIELPGFVAGNGHMAITK-----YDDVFYASRPDIFFSYSPNITINDQRPBELAEYF 107
+ : : : : : : : : : : : : : : : : : : : : : : : : : :
93 GEAYVDFPTSMDEPQKQFALANOVVGMPVDKLENRIQELACSLIESL---RPOGOCN 149
+ : : : : : : : : : : : : : : : : : : : : : : : : : :
108 GSM-----IYLDPRHQRLHSIVSRAFTPKVAVATEAIVADRAHRLVSSMIANNPDRQAD 162
+ : : : : : : : : : : : : : : : : : : : : : : : : : :
150 FTEDYAEPPPIRIFMLAGLPEEDIPHLKYLTDQMT---RPDGSMTFAEKKALYD---Y 203
+ : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
Db 163 LVSELAGPLQLQICDMGMIKPKADHRIFFMTNVILGEGPDLATQDFDEMOVSADIGAY 222
QY 204 LIPIDRORRPGCTDAISIVANGVNGRPTITSPKARMCGLLVGGLDVTYVNFSLFSMEF 263
Db 223 ATPLAERRRVNHHDDLSSLVEAEVDERSLSSREIASFFLLVAGNETRNATHGYLA 262
QY 264 LAKSPHEQELIERPERI -PACBELLRPS-LVADGRILITSDYEFHGVLKKGDOILLP 321
Db 283 LSHRPEQRDMWMSDFDLAETAVEEIVRMASPVYVMRRITLQDIELRGTKMAAGDKVSLW 342
QY 322 QMSGLDERENACPMHVDSPKRVSHHTF-GHSHCLGQHILARREIYVLKEMLTPIPD 380
Db 343 YCSANRDESFADPTFDLARNPNPHLGFGGGAHFCGLGANLARREIYVAFDELROMPD 402
QY 381 FSIAPGAQIOHKSGIVSGVQALPLW 406
Db 403 V-VATEEPARLLSOFHIGIKTLPVTW 427

RESULT 11
CPYK_BACSU
ID CPYK_BACSU STANDARD; PRT; 410 AA.
AC 008469;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE CYTOCHROME P450 (EC 1.14.-.-).
GN CYP4 OR CYP107J1.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=168 / BGSC1A1;
RX MEDLINE: 97431495
RA Belitsky B.R., Gustafsson M.C.U., Somenshein A.L., von Wachenfeldt C.;
RT "An lrp-like gene of Bacillus subtilis involved in branched-chain
RT amino acid transport."
RL J. Bacteriol. 179:5448-5457(1997).
RN RP
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE: 97453479.
RA Sorokin A., Bolotin A., Purnelle B., Hilbert H., Lauber J.;
RA Duesterhoeft A., Ehrlich S.D.;
RT "Sequence of the Bacillus subtilis genome region in the vicinity of
RT the lrp operon reveals two new extracytoplasmic function RNA
RT polymerase sigma factors Sigv and Sigz."
RL Microbiology 143:2939-2943(1997).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Y11043; CAAT1937.1; -
CC EMBL: 093876; AAB80896.1; -
CC EMBL: 299117; CAB14615.1; -
CC HSSP: Q00441; IOXA.
CC SUBTLIST: BG11929; CYP4.
CC PFM: PR00067; P450; 1.
CC PRINTS: PR00359; BP450.
CC PROSITE: PS00086; CYTOCHROME_P450; 1.
CC Oxidoreductase; Monooxygenase; Heme.
KW BINDING 359 359 HEME (BY SIMILARITY).
FT SEQUENCE 410 AA; 47384 MW; 035E98E58AA761AB CRC64;
```

Query Match 15.3%; Score 333; DB 1; Length 410;

```
Best local Similarity 27.2%; Pred. No. 3, 2e-18;
Matches 86; Conservative 69; Mismatches 139; Indels 22; Gaps 5;

QY 104 DPREGQFRLANQVGMPPVVDKLENRIQELACSLIESLRPOQOCNFTEDYAPFIRIF 163
Db 100 DPEDHNRKRLTVOKAFTTHRTITLOLEDKIOHTIADSLDKVOPKFMVLVDYAPPLIYI 159
QY 164 MLAGLPEEDIPMLKYLTDQMTRPDGSMTFAEKAEL-----VYLLPIIPORQ 213
Db 160 SEMGLPLEDROKFRWWSQAI-----IDFSDAPEFLQENDHLGFEVETLESTVAKKR 213
QY 214 KPGTDAISIVANGVNGRPTITSPKARMCGLLVGGLDVTYVNFSLFSMEFLAKSPHROE 273
Db 214 EPAGDILISALIQAESGTOI,STEEELYSMTILIVAGHEFTVNLITMWTYALMCHNDQLK 273
QY 274 LIRPERIPACBELLRPSLY--ADGRILITSDYEFHGVLKKGDOILLPOMLSGLDERE 331
Db 274 LRQOPDLMSAIEALRFHSPVELTTRWTAEPFLLHGEITRKDYITISLASANDERKI 333
QY 332 NACPMHVDSPKRVSHHTFGHSHCLGQHILARREIYVLKEMLTPIPPFSIAPGAQIOH 391
Db 334 FPNADFDIERKKNRHHIATFGHGHFCLGAQLARLEKIAISTLLRCPNIOI-KGEKKQM 392
QY 392 K---SGIVSGVQALPL 404
Db 393 KWKGNFLMRALIELPI 408

RESULT 12
CPYK_SACER
ID CPYK_SACER STANDARD; PRT; 405 AA.
AC P33271;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CYTOCHROME P450 107B1 (EC 1.14.-.-) (P450C11B1).
GN CYP107B1.
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Pseudonocardineae; Pseudonocardaceae;
OC Saccharopolyspora.
RN (1)
RP SEQUENCE FROM N.A., AND SEQUENCE OF 172-203 AND 302-324.
RC STRAIN=NRRL 2338;
RX MEDLINE: 92121109.
RA Andersen J.F., Hutchinson C.R.;
RT "Characterization of Saccharopolyspora erythraea cytochrome P-450
RT genes and enzymes, including 6-deoxyerythronolide B hydroxylase."
RL J. Bacteriol. 174:725-735(1992).
CC -1- FUNCTION: NOT KNOWN, PROBABLY INVOLVED IN THE CATABOLISM OF
CC OCTANE AND GUAIACOL. IT DISPLAYS A WEAK ACTIVITY IN THE O-
CC DEALKYLATION OF 7-ETHOXYCOUMARIN.
CC -----
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M83110; AAA26483.1; -
CC PIR: BA2606; B42606.
CC HSSP: Q00441; IOXA.
CC PFM: PR00067; P450; 1.
CC PRINTS: PR00359; BP450.
CC PROSITE: PS00086; CYTOCHROME_P450; 1.
CC Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.
KW BINDING 352 352 HEME (BY SIMILARITY).
FT SEQUENCE 405 AA; 45238 MW; 71C93CEC1FDC53FD CRC64;
```

Query Match 15.2%; Score 331.5; DB 1; Length 405;  
 Best Local Similarity 25.6%; Pred. No. 4.1e-18;  
 Matches 107; Conservative 73; Mismatches 185; Indels 53; Gaps 10;

18 VPEHLVFD---FDMYPSMLSGVQEAAMVLOESNVPDLVWTRCGN-GHWIATRGOLIR 72  
 6 VPDLLEAFDDAFADRRH-----RYARMEEPVQRI--RTVNGLDAMLITRYEDVK 53  
 73 EAEEDVRHFSSECPPIPREAGEAYDFIPTSM-----DPEEQOQFR 112  
 54 QALLD-----PRIAKDGGRTQOITIEKRLDAERRRGSPDLGPHMLNTPDPDHTRLR 105  
 113 ALANOVGMPEVDKLENRIGELACLSIESLRPOGOCNFTDYAEPPRIPIFELAGLPEE 172  
 106 KLVVKAFTARVAVGLRPRIGQITDDLLRLAGSEVDLIDEFAPRLPIITYISELMGVEDS 165  
 173 DIPHLKYLTDQMTPRDGSMTFAEAKR--ALYDYLPIIEGRQKPGTDAISIVANGOVNC 230  
 166 RRDDEFSGMTNVLV--DQSGPEAOQASVAMVEYTELIAKRTPEGDDLLTALLEAVEDG 223  
 231 RPIITDPAKMGCLLVGGLDYVNVNPLSPMEFLASPEHROELIEREPRIAPACEELR 290  
 224 DRISSEGLIAMVELLVAGHETTVNLGNCSLSLGNPDQALALRNDPSILPGATIEETLR 283  
 291 RFSLVADG--RLITSYEFHGVOLKKGQDILLPQMLSGLDERENACPMHVDERSOKVSH 348  
 284 YESPVANGTFRHRAEAVRFEDVVIPEBELVMAVLAANRGGERFEDPDRTDITRETTGHV 343  
 349 TFGHSHLCLGHLARREITVLTKEMLTIRIPDSIA--PGAIOHKSIGVSGVALPL 404  
 344 AFHGHIHFVGALARLEAOIAGVRLERPPDLMAASPDLLRMFRFSVLRLGLEKLEPV 401

RESULT 13  
 CPYE\_STRGO STANDARD; PRT; 405 AA.

AC P18326;  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE CYTOCHROME P450-SOI (EC 1.14.-.-) (P450-CVA1) (CYP105A1).  
 GN CYP105A1 OR SUVAC.  
 OS Streptomyces griseolus.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-31.  
 RC STRAIN=ATCC 11796;  
 RX Omer C.A., Lenstra R., Little P.J., Dean C., Tepperman J.M.,  
 RA Leto K.J., Komesser J.A., O'Keefe D.P.,  
 RT "Genes for two herbicide-inducible cytochromes P-450 from  
 Streptomyces griseolus."  
 RL J. Bacteriol. 172:3335-3345(1990).  
 CC -1- FUNCTION: METABOLISM OF A NUMBER OF SULFONYLUREA HERBICIDES.  
 CC -1- INDUCTION: BY HERBICIDE.  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: M32238; AAA26823.1; -  
 PIR: A35401; A35401.  
 DR HSSP: P23295; 2ROW.  
 DR PFAM: PF00067; P450; 1.  
 DR PRINTS: PR00359; BP450.  
 DR PROSITE: PS00086; CYTOCHROME\_P450; 1.  
 KM Oxidoreductase; Monooxygenase; Electron transport; Heme.

FT INIT\_MET 0 0  
 FT BINDING 354 354 HEME (BY SIMILARITY).  
 SQ SEQUENCE 405 AA; 44081 MW; 92AB36E064FD0B3E CRC64;

Query Match 15.1%; Score 330; DB 1; Length 405;  
 Best Local Similarity 26.3%; Pred. No. 5.4e-18;  
 Matches 104; Conservative 74; Mismatches 193; Indels 24; Gaps 11;

31 PSNLISAGVQ--EAMAVLOESNVPDLVWTRCGN-HWIATRGOLIREAVEDYR----- 79  
 16 PSNRSCPYOLPDGVYAOIRDTCPGLRHVTLVDYDQAWVYTKHKAARKLLGPRLSNRD 75  
 80 HESSECPPI--REAGAYVFIPTSMPPRQORFALANOVGMPEVDKLENIQELACS 137  
 76 NFPATSPFEAVRESPPAF--TGLDPRHGTTRRRRTISEFTVKRIKGRPREVEEVHG 131  
 138 LIESLRPOG--OCNFTDYAEPPRIPIFMLAGLPEEDIPHLKYLTDQMTPRDGSMTFAEA 196  
 132 FLDEMLAAGTADLVSGFALPVPSMYICRLGYPADHEFFQDASKRLVOSTDAQSLTA 191  
 197 KEALYDLPIIDQRRQKPGTDAI--SIVANGVNRPIITSDEAKRMCGLLVGLDTPVN 255  
 192 RNDLAGYLDLITQFQTEPGAGLVGALVADQLANG--IDREELISTAMLLIAGHETTAS 250  
 256 FLSPMEFLAKSPENRQELIEREPRIAPACEELRRSL--VADGRILTSDFEHGVOLK 313  
 251 MTSLSVTLTLDHPROYAALADRSLVGAVEELRLYLAINDIAGRAVATDIEVGHLLR 310  
 314 KGDQILLPQMLSGLDERENACPMHVDERSOKVSHTEFGHSHLCLGHLARREITVLKE 373  
 311 AGEGLIYVNSIANRDGVYEDPDALDIHRSARHHLAEGFVHQCLOGNLARLELEVILMA 370

RESULT 14  
 YC56\_MYCTU STANDARD; PRT; 405 AA.

AC Q11062;  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-FEB-2000 (Rel. 39, Last annotation update)  
 DE PUTATIVE CYTOCHROME P450 RV1256C (EC 1.14.-.-).  
 GN RV1256C OR MTCY50.26.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RX MEDLINE: 98295987.  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekcia F.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,  
 RA Horsbly T., Jagels K., Krogg A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skellern S., Squares S., Squires R., Sulston J.E.,  
 RA Taylor K., Whitehead S., Barrell B.G.,  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544(1998).  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

```
CC or send an email to license@sb-slb.ch).
CC -----
CC EMBL: 277137; CAB00896.1; -.
CC DR HSSP: P33006; ICPT.
CC DR TUBERCULIST: RV1256C; -.
CC DR PFAM: PF00067; P450; 2.
CC DR PROSITE: PS00086; CYTOCHROME_P450; 1.
CC KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.
CC FT BINDING 354 354 HEME (BY SIMILARITY).
CC SEQUENCE 405 AA; 44580 MW; 72DEA6CB688FA48 CRC64;

Query Match 14.7%; Score 320.5; DB 1; Length 405;
Best Local Similarity 29.5%; Pred. No. 2.8e-17;
Matches 102; Conservative 51; Mismatches 160; Indels 33; Gaps 9;

QY 62 HWIATGQLREAYEDYRHSSECPFIRREAGE-----ANDFIPTSM-DEPEQROFALA 115
DB 46 YVYLSRADVWSARHQFFSS-AQGLTVNYGLEMIGLHDPMPVMDPVHTEFKLV 104
QY 116 NQVYGVPRVYDKLENRIQELACSLIESLRPGQCNFTEDEYAEPPRIRIFMLAGLPEDIP 175
DB 105 SRFFTRQGVETVEPYRKFEYVERLEKIRANGGDIYTELEKPLPSMVVAHYLGVPEDWT 164
QY 176 HLRVLDQMTTP---DGSMTFA-EAKKALYDYLPIIEQRORPGTDAIS--IVANGVN 229
DB 165 QDFGWTQAIYAANAVDQATGALDAVGSMAVFTGLIERRTEPADDAISHVAAGVAD 224
QY 230 GRITTSDEAKRMGGLLVGLDVTNVNFLSMEFLAKSPHROELIERPRIPRPAACEELL 289
DB 225 GDTAGTLLSLAFTFTVYTGNDVTGMLGSMPLLNHRPQRLRLDDPGIDVAEELL 284
QY 290 RRRSLVAD-GRITTSDEYEFHGVLKKDQILLPQMLSGDERE-----NACPMHV 338
DB 285 RLSPVOGLARTTTRDTTIDTTPAGRRVLLYXGSANRBEROYGRDAELDYTRCPRTI 344
QY 339 DFRQKVSHTTFEGHSHLCQHLARREIIVTLKWLTRIPDFSIA 384
DB 345 -----LTFSGAHNHCIGAAARQCRAVLTTELLARCPDEVA 381

RESULT 15
BIOT_BACSU STANDARD; PRT; 395 AA.
ID BIOT_BACSU
AC P53554;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, last annotation update)
DT 15-FEB-2000 (Rel. 39, last annotation update)
DE BIOTIN BIOSYNTHESIS; CYTOCHROME P450-LIKE ENZYME (EC 1.14.-.-).
GN BIOT OR CYPT07H.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 96312354.
RA Bower S., Perkins J.B., Yocum R.R., Howitt C.L., Rahaim P.,
RA Pero J.;
RT "Cloning, sequencing, and characterization of the Bacillus subtilis
RT biotin biosynthetic operon.";
RL J. Bacteriol. 178:412-413(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 98048467.
RA Lapidus A., Galleron N., Sorokin A., Ehrlich S.D.;
RT "Sequencing and functional annotation of the Bacillus subtilis genes
RT in the 200 kb rmb-dnaB region.";
RL Microbiology 143:3431-3441(1997).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@sb-slb.ch).
CC -----
CC EMBL: U51868; AAB17462.1; -.
CC DR EMBL: AF008220; AAC00266.1; -.
CC DR EMBL: Z99119; CAB14997.1; -.
CC DR HSSP: Q00441; IOXA.
CC DR SUBTLIST: BG11528; BIOT.
CC DR PFAM: PF00067; P450; 1.
CC DR PRINTS: PR00359; BP450.
CC DR PROSITE: PS00086; CYTOCHROME_P450; 1.
CC KW Biotin biosynthesis; Oxidoreductase; Monooxygenase;
CC Heme electron transport; Heme.
CC FT BINDING 345 345 HEME (BY SIMILARITY).
CC SEQUENCE 395 AA; 44865 MW; E4AC3AF2637ACE1A CRC64;

Query Match 14.6%; Score 318.5; DB 1; Length 395;
Best Local Similarity 27.3%; Pred. No. 3.9e-17;
Matches 89; Conservative 65; Mismatches 133; Indels 39; Gaps 9;

QY 75 YEDYRHFSSSECPFIRREAGEAYDFIPTSMDEPQROFALANQVGMVYDKLENRIQEL 134
DB 70 YODLSHVQNMQLFQO-----PDHRLRLTLASGAFYPRPTTESYQPIIEF 115
QY 135 ACSLIESLRPGQCNFTEDEYAEPPRIRIFML--LADLPREDIPHLKLYLDQMTTPGSGMT 192
DB 116 VNHLLDQVQKKKMEVISPFA--FPLASPVIANIIGVPEDREOLKEMASLIQ--TID 170
QY 193 FAKKALYD-----YLPIIEQRORPGTDAISIVANGVNGRPITSDEAKRM 242
DB 171 FTSSRKALTEGNIMAVQAMAYFEKELQKRKRHQDOMISMLKGRKDK-LTEEEAASIC 229
QY 243 GLLVGLDVTNVNFLSMEFLAKSPHROELIERPRIPRPAACEELLRRPS-LVADGRIL 301
DB 230 ILAIAIGHETVNLISNSVLCILQHPQOLKLENDPLIGTAVEECIAYESPOTMARVA 289
QY 302 TSDYEFHGVLKKDQILLPQMLSGDERENMC---PMHVPFRQKVSHTTFEGHSHLC 358
DB 290 SEDIDICGVTIKQGEQVYL---LLGANNDPSTLFTNPVDVITRSPNHLRSGHHVCL 346
QY 359 GQHLARREIIVTLKWLTRIPDFSIA 384
DB 347 GSSLARLEAQIAINTLLQRRPISLNTLA 372
```

Search completed: October 4, 2000, 13:04:51  
Job time: 1682 sec





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 06:43:07 ; Search time 79.83 Seconds

(without alignments)  
359,568 Million cell updates/sec

Title: US-09-246-451-2

Perfect score: 2180

Sequence: 1 TTEMIOGNANLAPLPHPHVE.....IVSGVQALPLVWDPATTKAV 414

Scoring table: BLOSUM62

Searched: 225878 segs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: SPREMBL\_12:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mammal:\*  
8: sp\_mhc:\*  
9: sp\_organellar:\*  
10: sp\_phase:\*  
11: sp\_plant:\*  
12: sp\_rodent:\*  
13: sp\_virus:\*  
14: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	365	16.7	2	059523	059523 micromonospora
2	363.5	16.7	2	087605	087605 streptomyc
3	360	16.5	2	052544	052544 amycolatops
4	346.5	15.9	2	054302	054302 streptomyc
5	346	15.9	2	09X5P9	09X5P9 streptomyc
6	344.5	15.8	2	059819	059819 streptomyc
7	329	15.1	2	032460	032460 actinomyc
8	326	15.0	2	059723	059723 pseudomonas
9	325.5	14.9	2	092H01	092H01 streptomyc
10	324.5	14.9	2	031785	031785 bacillus su
11	321	14.7	2	09X5P8	09X5P8 streptomyc
12	318	14.6	2	032927	032927 mycobacteri
13	316	14.5	2	09X9P7	09X9P7 streptomyc
14	313.5	14.4	2	059831	059831 streptomyc
15	313	14.4	2	087192	087192 streptomyc
16	312	14.3	2	09XDB0	09XDB0 mycobacteri
17	304	13.9	2	059910	059910 streptomyc
18	301	13.8	2	085697	085697 streptomyc
19	299.5	13.7	2	024727	024727 nocardioide

20	294	13.5	351	2	052572	052572 amycolatops
21	293.5	13.5	337	2	052816	052816 amycolatops
22	291	13.3	388	2	P77977	P77977 streptomyc
23	290.5	13.3	406	2	087675	087675 amycolatops
24	276.5	12.7	310	2	P96562	P96562 amycolatops
25	275	12.6	421	2	052561	052561 amycolatops
26	267	12.2	398	2	087674	087674 amycolatops
27	265.5	12.2	411	2	09X803	09X803 streptomyc
28	264	12.1	403	2	052560	052560 amycolatops
29	263.5	12.1	437	2	09ZFC0	09ZFC0 mycobacteri
30	261.5	12.0	386	2	059921	059921 streptomyc
31	261	12.0	395	2	09X5T2	09X5T2 streptomyc
32	254	11.7	406	2	052823	052823 amycolatops
33	254	11.7	426	2	09ZFC3	09ZFC3 mycobacteri
34	248	11.4	411	2	060005	060005 s putative
35	244.5	11.2	511	2	052569	052569 amycolatops
36	240	11.0	433	2	09XCC6	09XCC6 streptomyc
37	225	10.3	391	2	052822	052822 amycolatops
38	225	10.3	398	2	059079	059079 amycolata a
39	220.5	10.1	391	2	087673	087673 amycolatops
40	217	10.0	397	2	052802	052802 agrobacteri
41	204.5	9.4	419	2	050242	050242 mycobacteri
42	182.5	8.4	119	2	085655	085655 streptomyc
43	175	8.0	120	2	085653	085653 streptococ
44	174.5	8.0	313	2	09X418	09X418 myxococcus
45	165	7.6	120	2	085650	085650 streptomyc

## ALIGNMENTS

RESULT	ID	PRELIMINARY	PRT	397 AA.
059523	059523	059523		
AC	059523	059523		
DT	01-NOV-1996 (TREMBLrel. 01, Created)			
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)			
DT	01-NOV-1999 (TREMBLrel. 12, Last annotation update)			
DE	CYTOSOLIC P450-LIKE PROTEIN (EC 1.14.14.1).			
GN	MYC.			
OS	Micromonospora griseorubida.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
OC	Actinomycetales; Micromonosporinae; Micromonosporaceae;			
OC	Micromonospora.			
RN	[1]			
RC	SEQUENCE FROM N.A.			
RC	STRAIN-A11725;			
RX	MDLINE; 95107242.			
RA	INOUE M., TAKADA Y., MUTO N., HORINOUCI S., BEPPU T.;			
RT	"Characterization and expression of a p-450-like mycinamicin			
RT	biosynthesis gene using a novel Micromonospora-Escherichia coli			
RT	shuttle cosmid vector".			
RL	Mol. Gen. Genet. 245:456-464(1994).			
DR	EMBL; D16098; BAA03672.1; .			
DR	HSSP; O00441; IOXA.			
DR	PRAM; PFO0067; P450; 1.			
DR	PRINTS; PR00359; BP450.			
SQ	SEQUENCE 397 AA; 44331 MW; D1D592A6 CRC32;			
Query Match	16.7%; Score 365; DB 2; Length 397;			
Best local Similarity	28.5%; Pred. No. 7.2e-22;			
Matches	105; Conservative 57; Mismatches 168; Indels 38; Gaps 8;			
OY	63 WIATRGILREAYDYHFSSECF-----PREAGEAYFIPTSMPPROFRA 113			
Db	44 WLVR-----YEDVRAVLGDGRFVGPSWTRDEPRTPRMVVGGLISMPDPSRLR 96			
OY	114 LANVGMPVVDKENTRIOLACSLIESLRPGO-CNFTEDYAPFPRIFFMLAGPDE 172			
Db	97 LVYKAFTRARRASLRPARLAHELVDOMATGPPADLVAMFAQLPVRVYICELGVPSA 156			
OY	173 DIPHLKYLTDQMTFRDQSM-----TFAEAKALYDYLPIIEORORRPGTDAISIV 223			

```

Db 157 D-----HRTFRWGSAGLSTAEVTAENQEAQYAAVWGDLIRRRKRPDDDLVSAL 209
OY 224 ANGVNAGRPTTSDAKRMCGLLVGLDVTYVNFSLFSMEFLAKSPEROLIERPERIPA 283
Db 210 VQARDDODSDISBELDLALGGLVAGYESTTTOIADFYVLLMTPELRKOLLDPPELIPS 269
OY 284 ACEELLARRESL---VADGRILTSDFEFGVQAKKQDQILLPOMLSGLDERENACPMHVP 340
Db 270 AVELLTRWVPLGVTAFPRYAVEDVTLRGVTRIRAGEPVLASTGANNRDAQFPDADRIDV 329
OY 341 SRKAVHTTTFGHSILCLGHLARRELYTLKEMLRIRIPFSLA-PCAQIQHSG-IVSG 386
Db 330 DRPNQHLGFGHGHCLAPLARVELQVALLEVLQRLPGIRLGPETQLRWSEGMILRG 389
OY 399 VQALPLVW 406
Db 390 PLEPLVW 397

```

## RESULT 2

```

OY 087605 PRELIMINARY: PRT: 416 AA.
AC 087605:
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, last sequence update)
DE CYTOCHROME P450 MONOOXYGENASE.
DE PICK OR PICK.
OS Streptomyces violaceus (Streptomyces venezuelae), and
OS Streptomyces venezuelae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC15439;
RA BETLACH M.C., KEALEY J.T., BETLACH M.C., ASHLEY G.W., MCDANIEL R.;
RT "Characterization of the macroliide P450 hydroxylase from Streptomyces
RT venezuelae which converts narbomycin to picromycin.";
RL Biochemistry 0:0-0(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES-S.VENEZUELA; STRAIN=ATCC15439;
RX MEDLINE: 99051447.
RA XUE Y., WILSON D., ZHAO L., LIU H.-W., SHERMAN D.H.;
RT "Hydroxylation of macroliolones YC-17 and narbomycin is mediated by
RT the pikC-encoded cytochrome P450 in Streptomyces venezuelae.";
RL Chem. Biol. 5:661-669(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES-S.VENEZUELA; STRAIN=ATCC15439;
RX MEDLINE: 9844533.
RA XUE Y., ZHAO L., LIU H.-W., SHERMAN D.H.;
RT "A gene cluster for macroliide antibiotic biosynthesis in Streptomyces
RT venezuelae: architecture of metabolic diversity.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:12111-12116(1998).
DR EMBL: AF079139; AAC6886.1; -.
DR EMBL: AF079139; AAC6886.1; -.
DR HSSP: 000441; 10XA.
DR PFAM: PF00067; P450; 1.
DR PRINTS: PR00359; BP450.
KW Monooxygenase.
SQ SEQUENCE 416 AA: 46038 MW; CAD6EBA0 CRC32;

```

Query Match 16.7%; Score 363.5; DB 2; Length 416;  
 Best Local Similarity 28.3%; Pred. No. 1e-21;  
 Matches 97; Conservative 61; Mismatches 164; Indels 21; Gaps 7;

OY 76 EDYRHSSSCPTIPRAGAYDPIPTSMDEQORFALANQVGVVVDKLENRTIOELA 135  
 Db 72 KWR--NSTTPTLEADALNHNMLS--DPRHTRRLKLVAREETKRWELLKRPVOELV 127

```

OY 136 CSLIESL--RPGQCNFTEDYAEFPPIRLFMILLAGLPREDIPHLKYLTDQMPRDSMTF 193
Db 128 DGLVDMALAPDGRDLMEIAMPRLPTTVISELLGVEPDRAPAFVWDTAFVDRPPQA 187
OY 194 AEAKKALYDLIPIIDORRCKPCTDAIS-IVANGVNRGRTTDEAKRMCGLLVGLDT 252
Db 188 QTAEMMSGYLSRLIDSKRQDDEDLALVRSDEGSRLTSEELGMANHLLVAGHET 247
OY 253 VVNFSLFSMEFLAKSPEROLIERPERIPACCELLRRSLVADRIILTSDEF----- 307
Db 248 TVNLIANGMTALLSHPDQALALRADMTILDGAVEMLR-----YEGVESATIRFVEPY 302
OY 308 --HGVQKKQDQILLPOMLSGLDERENACPMHVPFSROKVSHTTFGSHILGOMIARR 365
Db 303 DLDGTVIPAGDVTLYVLAADHRTPERPDRPHRDTRDAGHLAFGHIFCIGAPLARL 362
OY 366 EITVTLKEMLRIPDS--IAPGAQIQHSGIVSQVALPLVW 406
Db 363 EARIAVALLERCPDLALDVSPEGELWYPMPIRGKALPIRW 405

```

## RESULT 3

```

OY 052544 PRELIMINARY: PRT: 396 AA.
AC 052544:
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, last sequence update)
DE CYTOCHROME P450 MONOOXYGENASE.
DE Amycolatopsis mediterranei.
OS Amycolatopsis mediterranei.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Pseudonocardiales; Pseudonocardaceae; Amycolatopsis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S699;
RA KIM C.G., YU T.W., FRYHLE C., HANDA S., FLOSS H.G.;
RL J. Biol. Chem. 0:0-0(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S699;
RA AUGUST P.R., TANG L., YOON Y.J., NING S., MUELLER R., YU T.W.,
RA TAYLOR M., HOFFMANN D., KIM C.G., ZHANG X., HUTCHINSON C.R.,
RA FLOSS H.G.;
RL Chem. Biol. 5:0-0(0002).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S699;
RA AUGUST P.R., TANG L., YOON Y.J., NING S., MUELLER R., HUTCHINSON C.R.,
RA FLOSS H.G.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF040570; AAC01709.1; -.
DR HSSP: 000441; 10XA.
DR PFAM: PF00067; P450; 1.
DR PRINTS: PR00359; BP450.
KW Monooxygenase.
SQ SEQUENCE 396 AA: 44293 MW; BA875129 CRC32;

```

Query Match 16.5%; Score 360; DB 2; Length 396;  
 Best Local Similarity 28.8%; Pred. No. 1.8e-21;  
 Matches 116; Conservative 60; Mismatches 183; Indels 44; Gaps 14;

OY 27 DWNPNLSAGVOEMAVYLOESNVPLVTRNGGH--WIAIRGQILREAYEDYRFSSE 84  
 Db 15 DKFDPVAVFDSLRER-RPLAKMYPD-----GHVGIVSSYELVREVLSDLR-FSHS 64  
OY 85 CP---FIPREAGEAVDFIPT-----SMDPEQORFALANQVGVVVDKLENRTIO 132  
Db 65 CEVGHFPPVTHQGV--IFTHPLIPGMFTHMDPEHTRRKLLTGFTYRRASRLIPAE 121  
OY 133 ELACSLIESLRPG-OCNFTEDYAEFPPIRIFMLAGLPED---IPLKYLTDQMPR 187  
Db 122 AVAAEOIEVWRAKGAADVVMDFAKPLVRLMGLGELVGLPEERDRVPAVTLHLDAEAP 181

QY	188	DGMYFAKAKALDYLLPIIEORRQKPGDASIVANGQVNGRPITSDAKRMCGLLV	247
Db	182	AEA---AAAYVACKFDEYVERERRRQRDDLSISLV- ---EDLQEBLRTNVTLLF	233
QY	248	GLDPTVVNFISFSMEFLAKSPENHOELIERPERIPACCELLRRFSL- -VADGRILTSY	305
Db	234	AGYETTELKALATGVFALLNHTDQALALRAEREKIDAAIIEBLRKLTVYNGHYRTALEDV	293
QY	306	EFHGVALTKGQDOLLPMQLSGLDERENACPMHVDFSQKVSHHTFGHSNLCGHLARR	365
Db	294	KLEBELLKKGTVTVSLPAAARNRDPAKFGCPALIEDTSGHVAFGGIGHQICGONLARI	353
QY	366	EIYTLKEMLRIPDESLA- PGAOIQHK- SCIVSGVALPLVM	406
Db	354	ELRAGFTALLRAFPBLRLVAPADVPRLRLKQSVSVYKLLPLSV	396

RESULT	4			
054302		PRELIMINARY;	PRT;	404 AA.
ID	Q54302			
AC	Q54302;			
DT	01-NOV-1996 (TrEMBLrel. 01, Created)			
DT	01-NOV-1996 (TrEMBLrel. 01, last sequence update)			
DT	01-NOV-1999 (TrEMBLrel. 12, last annotation update)			
DE	CTROCHROME P450.			
GN	RAPN.			
OS	Streptomyces hygroscopicus.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
OC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=NRRL 5491;			
RX	MEDLINE: 95372374.			
RA	SCHWECKE T., APARICIO J.F., MOLNAR I., KOENIG A., KHAW L.E.,			
RA	HAYDOCK S.F., OLITNYK M., CAFFEY P., CORTES J., LESTER J.B.,			
RA	BOEHM G.A., STAUNTON J., LEADLAY P.F.;			
RT	"The biosynthetic gene cluster for the polyketide immunosuppressant			
RT	rapamycin.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 92:7839-7843(1995).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=NRRL 5491;			
RA	MOLNAR I., APARICIO J.F., HAYDOCK S.F., EE KHAW L., SCHWECKE T.,			
RA	KOENIG A., STAUNTON J., LEADLAY P.F., STAUNTON J., LEADLAY P.F.;			
RL	Gene 0:0-0(0).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=NRRL 5491;			
RA	APARICIO J.F., MOLNAR I., SCHWECKE T., KOENIG A., HAYDOCK S.F.,			
RA	EE KHAW L., STAUNTON J., LEADLAY F., LESTER J.B., BOEHM G.A.,			
RA	STAUNTON J., LEADLAY P.F.;			
RL	Gene 0:0-0(0).			
DR	EMBL: X86780; CAA60465.1; -			
DR	HSSP: X00441; I0XA			
DR	PRAM: PF00067; P450; 1.			
DR	PRINTS: PR00359; BP450.			
DR	SEQUENCE 404 AA; 45071 MW; 05AB94DF CRC32;			

	Query Match	15.9%	Score 346.5;	DB 2;	Length 404;	
	Best Local Similarity	28.6%;	Pred. No.2.3e+20;			
	Matches 102; Conservative	63;	Mismatches 177;	Indels 15;	Gaps 9	
Qy	63	WIATRGOLIREAYEDYRRHFSSECPF---IPRAGEAYDF-IPTSMDPPEQRGRALANOV	118			
		:   : :           :   :   :   :   :   :   :   :				
Db	50	WLAASWEVAKVFVDPR-FSRSATIGKDVPRPLAIQDPVTMLMDPPETHRLRRATVA	108			
		:   :   :   :   :   :   :   :   :   :   :   :   :				
Qy	119	VGMVPVADKLERNIOELACSLIESLRPG-QCVFTEDYAEPPRIFRIMLAGLPEEDIPIH	177			
		: : :   :   :   :   :   :   :   :   :   :   :   :   :				
Db	109	LTSRMELRPRTGEOVADDIDIKMLAKAGAPADMEDFALPETITMCELLGYPFIEQGTAF	168			
		:   :   :   :   :   :   :   :   :   :   :   :   :				
Qy	178	KYLTDQMTRPDGSMT--FAEAREALDYDLPILEORROKPGCTDAISIVANGOVNCRPTT	234			
		:   :   :   :   :   :   :   :   :   :   :   :   :				

[illegible]

RESULT	5		
09X5P9			
ID	09X5P9	PRELIMINARY;	PRT; 410 AA.
AC	09X5P9;		
DT	01-NOV-1999 (TREMBLrel. 12, Created)		
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)		
DT	01-NOV-1999 (TREMBLrel. 12, Last annotation update)		
DE	CYCROCHROME P450 HYDROXYLASE ORF3.		
OS	<i>Streptomyces lavendulae</i> .		
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;		
OC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RP	SPRAIN-NRRL 2564.		
RC	MEDLINE; 99201491.		
RX	MAO Y.Q., VAHOGU M., SHERMAN D.H.;		
RA	"Molecular characterization and analysis of the biosynthetic gene		
RT	cluster for the antitumor antibiotic mitomycin C from <i>Streptomyces</i>		
RT	<i>lavendulae</i> NRRL 2564.";		
RL	Chem. Biol. 6:251-263(1999).		
DR	EMBL; AF17374; AAD8449.1; -		
SO	SEQUENCE 410 AA; 45887 MW; 507C7F38 CRC32;		

Query Match 113: 15.9%; Score 346; DB 2; Length 410; Best Local Similarity 27.0%; Pred. No. 2.6e-20; Matches 113; Conservative 69; Mismatches 19; Indels 44; Gaps 14;

```
RESULT 6
ID 059819 PRELIMINARY; PRT; 407 AA.
AC 059819;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
DE 01-NOV-1999 (TREMblrel. 12, last annotation update)
DE CYTOCHROME P450 (EC 1.14.14.1).
GN OLEP.
OS Streptomyces antibioticus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
  Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95255619.
RA RODRIGUEZ A.M., OLANO C., MENDEZ C., HUTCHINSON C.R., SALAS J.A.;
  "A cytochrome P450-like gene possibly involved in oleandomycin
  biosynthesis by Streptomyces antibioticus.";
  FEMS Microbiol. Lett. 127:117-120(1995).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; U37200; AAA92553.1; -.
DR HSSP; 000441; 10XA.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
DR PFAM; PF00067; P450; 1.
DR PRINTS; PR00359; BP450.
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.
FT BINDING 356 HEME (BY SIMILARITY).
SQ SEQUENCE 407 AA; 44957 MW; 04411C60 CRC32;
```

```
Query Match 15.8%; Score 344.5; DB 2; Length 407;
Best Local Similarity 28.1%; Pred. No. 3.4e-20;
Matches 104; Conservative 63; Mismatches 168; Indels 35; Gaps 11;
```

```
QY 60 GGHWITRQOLIREAVEDYEDYRHSSEC---PTIPREAGEAVDFIPT-----SMDPEPQ 108
DB 50 GRALMTTRMSDARIIVLGDSR-FSTAAATDPATPR-----MFTPEPDGVLADDPDH 101
QY 109 RQFRALANOVGMVVDKLENRIQELACSLIESLRPOGQ-CNFTEDYAEFPPIRIMFLA 167
DB 102 TRRLRLVGKAFTRARVEEMRPVRSLVDSDMAHSPADLVEFLAVFPVAVICELL 161
QY 168 GLPEEDIPHLKYITDQM---TRPDGSMTPAEAKALYDYL---PIIEORROKPGDAIS 221
DB 162 GVALEDRDLEFRFESDAMLSTR---LTAAEIQRYOODEFVYMDGLVAGRDAPTEDLIG 217
QY 222 IVANGQVNGRPITSDAKRMCGLLLVGGLDVTYVNFLSFSMEFLAKSPHROELIERPERI 281
DB 218 ALALATNDNDHLTKGEIVNMGVSLTAGHETSYNQITNLVHLLTERKRRRESLVADPALV 277
QY 282 PACEELLRRFSVLADG---RILTSYEFHGVOLKKGDQILLPQMLSGLDERENACPMHY 338
DB 278 PAAVEEMLRTPLCVASGSRVATEDELSTLYVRAGECPVYHFAANDDEVEFDIADDEL 337
QY 339 DESROKVSHTTEGSHLCLGOHLARREITVTLEKWLTRIPDSIA-PSAQIQHKSQ-IV 396
DB 338 DFHREHNPPIAHFGHAGHICGAIQLGRLEQELALSALVREPTLDIAEPYAGLKWKGMI 397
QY 397 SGVQALPLWV 406
DB 398 RGLEROIVSM 407
```

```
RESULT 7
ID 032460 PRELIMINARY; PRT; 411 AA.
AC 032460;
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JAN-1998 (TREMblrel. 05, last sequence update)
DE 01-NOV-1999 (TREMblrel. 12, last annotation update)
DE ORF 10.
OS Actinomadura hibisica.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
```

```
OC Actinomycetales; Streptosporangineae; Thermomonosporaceae;
  Actinomadura.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-P157-2.
RA DAIRI T., HAMANO Y., IGARASHI Y., FURUMAI T., OKI T.;
  "Cloning and nucleotide sequence of the putative polyketide synthase
  genes for pradiacin biosynthesis from Actinomadura hibisica.";
  Biosci. Biotechnol. Biochem. 61:1445-1453(1997).
DR EMBL; D87924; BAA23153.1; -.
DR HSSP; 000441; 10XA.
DR PFAM; PF00067; P450; 1.
DR PRINTS; PR00359; BP450.
SQ SEQUENCE 411 AA; 44860 MW; 98A628B6 CRC32;
```

```
Query Match 15.1%; Score 329; DB 2; Length 411;
Best Local Similarity 26.4%; Pred. No. 6.3e-19;
Matches 114; Conservative 61; Mismatches 199; Indels 58; Gaps 16;
```

```
QY 4 TIOSNNLPLRPHVPEHLVFPEDMYNP-----SNLSAGVOEAMAVQESNVDPDVTWTRCN 59
DB 9 TVDPRPDVTPAFPPRPD-----DPQPCERARLSPDPVAKVLP-----T 50
QY 60 GGH-WIATGQOLIREAVEDYEDYRHSSEC---PTIPREAGEAVDFIPT-----MDPEP 107
DB 51 GDHAWVYTRADYRFVTSQ-RRSKKAAYVRPGAPR-----LIPQROGSKSLVIMDPPE 102
QY 108 QROFRALANOVGMVVDKLENRIQELACSLIESLRPOGQ-CNFTEDYAEFPPIRIMFL 166
DB 103 HFRMRKIVSRAFTARVEEMRAHVRDLTSGFVDEHVGPPADLIHLALPLPVTVICEM 162
QY 167 AGPEEDIPHLKYITDQMTRPDGSMTPAEAKE-----ALYDYLPIIIOORQKPGTDAI 220
DB 163 LGVPPDRPRFQDWTDRML-TTGAPALQADETKAAVGRKGLAELIDAKTAAPRADLL 221
QY 221 STVANGQVNGRPITSDAKRMCGLLLVGGLDVTYVNFLSFSMEFLAKSPHROELIERPER 280
DB 222 SLSTRHADQ-GLSEELLTFGKTLAAGYHTTAITHSYHILREPSRYARLRDRDPG 280
QY 281 IPACEELLRRFSVLADG---RILTSYEFHGVOLKKGDQILLPQMLSGLDERENACPM 336
DB 281 IPAVAEELL-RYGOIGGAGAIRIAYEDVEVGCTIVRAGEAVPLPFPANRPEVADPE 339
QY 337 HDFSROKVSHTTEGSHLCLGOHLARREITVTLEKWLTRIPDSIA-PSAQIQHKSQ 395
DB 340 ELDUGRTDNPPIAHGHIYCLGAPLARLELOVLETVERPALRLAIDADITWRPGL 399
QY 396 V-SGVQALPLWV 406
DB 400 AFARPDALPIAW 411
```

```
RESULT 8
ID 059723 PRELIMINARY; PRT; 406 AA.
AC 059723;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
DE 01-NOV-1999 (TREMblrel. 12, last annotation update)
DE CYTOCHROME P450 LIN (EC 1.14.14.1).
GN LINC.
OS pseudomonas incognita.
OC Bacteria; Proteobacteria.
RN [1]
RP SEQUENCE FROM N.A.
RA ROPP J.D., GUNSAIUS I.C., SLIGAR S.G.;
  Submitted (JAN-1994) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; U23310; AAA25810.1; -.
DR HSSP; P33006; 1CPT.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
```







Db 240 PDEPDPSPSSRHIAFAVGSFHICGALARLEAVTILSAISARPOVOLA-GELVYKPNV 298  
 QY 395 IVSGVALPL 404  
 Db 299 AMRGMALPV 308

## RESULT 13

QY9X9P7 PRELIMINARY; PRT; 410 AA.

AC QY9X9P7;  
 DT 01-NOV-1999 (Tremblrel. 12, Created)  
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
 DE NIKF PROTEIN.  
 GN NIKF.  
 OS Streptomyces tendae.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-TUE901;  
 RA BRUNNER C., LAUER B., SCHWARZ W., MOERLE V., BORMANN C.;  
 RT "Molecular characterization of co-transcribed genes from Streptomyces tendae Tu901 involved in the biosynthesis of the peptidyl moiety of the peptidyl nucleoside antibiotic nikkomycin."  
 RL Mol. Gen. Genet. 0:0-0(0).  
 DR EMBL: Y18574; CAB46536.1; --  
 SO SEQUENCE 410 AA; 45884 MW; 9B52CB74 CRC32;

Query Match 14.5%; Score 316; DB 2; Length 410;  
 Best Local Similarity 27.1%; Pred. No. 7.2e-18;

Matches 98; Conservative 64; Mismatches 180; Indels 20; Gaps 9;

QY 63 WIATRGQILREAYEDYRHFSSEC--PFI-----REAGEAYDFIPI-----SMDPEQR 109  
 Db 47 WLVKHLAKRLADPRVSDRLHRAFRRLTAQRTATEVRRLSTRSMIHLDGSEHG 106  
 QY 110 QFRALANQVGMRYVDKLENRIQELA-CSLIESLRPGQCNFTEDVAERPPRIEMLLAG 168  
 Db 107 AHKKILGEFSRLRIALRPRVQEIYDRSIDEMLAQPADLVHVSQAVPSLVICELG 166  
 QY 169 LRPEDIRHL-KYLLDQTRRDGSMTEFAEKAELYDILPII-EGRORCPGDAISIYANG 226  
 Db 167 VPIHQRDHFHEMGMLVSRVSIRERAASDADLDFLEDVTEKEKEPDDLLGRLLAR 226  
 QY 227 QVNGRPITSDPAKRMGCLLVGLDVTNVNPLFSMEFLAKSPENROELIERPERIPACE 286  
 Db 227 NRKRPVTHDEIVGTAVMLLAGHQTANNMISLGVALLENPEHKARIADPSLLPAIE 286  
 QY 287 ELARRFSLV--ADGRILTSDYEFHQVQKKGDQILLPOMLSGLDERENACPMHVDFSROK 344  
 Db 287 EMRLYFESVENAPARVATEDIEIGVITIRKDEGIVSGLAQMDDEVFHHDRIDFERGA 346  
 QY 345 VSHITFGHSHLCLGHLARREITVTKEMLTRIPDSIA-PGAQIQHKSQT-VSGVAL 402  
 Db 347 RHNVAFGYGHOCIGOMLAVLEIVETLLRRVPGSLAVPAEELPYKDDAGIYIYRV 406  
 QY 403 PL 404  
 Db 407 PV 408

## RESULT 14

QY9831 PRELIMINARY; PRT; 410 AA.

AC QY9831;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
 DE CYTOCHROME P450 SCA-2 (EC 1.14.14.1).

OS Streptomyces carbophilus.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN-SANK 62585;  
 RX MEDLINE: 96001248.  
 RA WATANABE I., NARA F., SERIZAWA N.;  
 RT "Cloning, characterization and expression of the gene encoding cytochrome P-450sca-2 from Streptomyces carbophilus involved in production of pravastatin, a specific HMG-CoA reductase inhibitor."  
 RL Gene 163:81-85(1995).  
 DR EMBL: D30815; BAA06492.1; --  
 DR HSSP: 000441; 10XA.  
 DR PFAM: PF00067; P450; 1.  
 DR PRINTS: PR00359; BP450.  
 SO SEQUENCE 410 AA; 45049 MW; 2DF70C8E CRC32;

Query Match 14.4%; Score 313.5; DB 2; Length 410;  
 Best Local Similarity 25.6%; Pred. No. 1.1e-17;  
 Matches 92; Conservative 63; Mismatches 185; Indels 19; Gaps 7;

QY 63 WIATRGQILREAYEDYR-----HFSSECPIPI--REAGEAYDFIPTSMDPEQRQFR 112  
 Db 56 WVVYKHEAARLLADPRLSRDRLADPRATSPRKARQSGPAR----IGMDPEHSTR 111  
 QY 113 ALANQVGMRYVDKLENRIQELACSLIESLRPG-QCNFTEDVAERPPRIEMLLAGLE 171  
 Db 112 RMTISEFTYKRIKGMRPDVERIVHGFIIDMLAAGPTADLVSQFALPVPMSVICHMLGVRY 171  
 QY 172 EDIRPLKYLIDQMPRPDSMTFAEKAELYDILPIIEGRORCPGDAISIYANGVNGR 231  
 Db 172 ADHEFFDASKRLVQAVDASAVARDFERKYLDTKLESEPGTLGKLVTHQADG 231  
 QY 232 PITSDEAKRMGCLLVGLDVTNVNPLFSMEFLAKSPENROELIERPERIPACEELLRR 291  
 Db 232 EIDRAELISTALLLVAGHETASMTSLVTLLEHNDQNALRADSLVPGANVEELLRV 291  
 QY 292 ESL--VADGRILTSYDFEHQVQKKGDQILLPOMLSGLDERENACPMHVDFSROKVSHTT 349  
 Db 292 LAIDAGIGRIATADITDQDLIRAGEGVITVNSIANRDSVFENPRDLRVHNRHNL 351  
 QY 350 FGHSHLCLGHLARREITVTKEMLTRIPDSIA-PGAQIQHKSQT-IVSGVALPLVW 406  
 Db 352 FGIVHQCCLQMLARLEIVLTVLEDRIPTRLAVVEEDLTLRPGTIGVNELPVTW 410

## RESULT 15

QY87192 PRELIMINARY; PRT; 406 AA.

AC QY87192;  
 DT 01-NOV-1998 (Tremblrel. 08, Created)  
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)  
 DE ONE-2 PROTEIN.  
 OS Streptomyces griseus.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA UEDA K.;  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 RX MEDLINE: 95378086.  
 RA UEDA K., KIM K.M., BEPPU T., HORINOCHI S.;  
 RT "Overexpression of a gene cluster encoding a chalcone synthase-like protein confers reddish brown pigment production in Streptomyces griseus."  
 RL J. Antibiot. 48:638-646(1995).  
 DR EMBL: AB018074; BAA33494.1; --  
 DR HSSP: 000441; 10XA.  
 DR PFAM: PF00067; P450; 1.



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 12:35:13 ; Search time 76.18 seconds

(without alignments)  
128.722 Million cell updates/sec

Title: US-09-246-451-11

Perfect score: 2180

Sequence: 1 TFEETIOSNANIKAPLPHPVE.....IVSGVOALPLWDPATTKAV 414

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: A\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2133	97.8	405	1 Y04128	Pseudomonas putida
2	1185	54.4	446	1 Y04126	Bacterial and mamm
3	364	16.7	396	1 R60777	Mycimycin IV hyd
4	332	15.2	408	1 R77867	S. clavuligerus OR
5	330.5	15.2	587	1 W33274	S. fradiae tylosin
6	330	15.1	406	1 R11349	Cytochrome enzyme
7	329	15.1	411	1 W54389	Actinomadura hibis
8	324.5	14.9	398	1 W11585	Streptomyces prist
9	322	14.8	403	1 R11350	Cytochrome enzyme
10	316	14.5	412	1 R38309	Sequence of the P4
11	313.5	14.4	410	1 R51368	Protein containing
12	296.5	13.6	404	1 R14724	6-hydroxylase enco
13	190	8.7	398	1 R47521	Vitamin D hydroxyl
14	160	7.3	422	1 W36128	Dauromycin C-14 hy
15	160	7.3	474	1 W36132	N-terminal modifie
16	155	7.1	422	1 W00729	Dauromycin 14-hy
17	144	6.6	518	1 W67616	A. nidulans phenyl
18	144	6.6	533	1 R15057	Cytochrome P450C25
19	142	6.5	494	1 R62825	Human steroid-21-h
20	130	6.0	512	1 W93216	Human cytochrome P
21	128	5.9	508	1 W35711	Chrysanthum flavon
22	128	5.9	512	1 R72365	Human auxillary cy
23	128	5.9	512	1 R33172	Human cytochrome P
24	128	5.9	516	1 W67617	P. chrysogenum phe
25	127.5	5.8	1169	1 R76544	Mitochondrial cyto
26	127	5.8	512	1 R72366	Human auxillary cy
27	127	5.8	512	1 W00652	Cytochrome P450A1
28	127	5.8	512	1 R33173	Human cytochrome P
29	125	5.7	524	1 P70577	Rat hepato-cytochr
30	122.5	5.6	523	1 R59291	Rat liver cytochro
31	122.5	5.6	898	1 P61030	Entire coded seque
32	122.5	5.6	898	1 P61082	Complete translati
33	122.5	5.6	899	1 P61056	Translation of pla

34	122.5	5.6	1144	1 P81334	Expression prod. o
35	122.5	5.6	1150	1 P81335	Expression prod. o
36	122.5	5.6	1150	1 P81337	Expression prod. o
37	122.5	5.6	1162	1 P81336	Expression prod. o
38	122	5.6	503	1 R72363	Human cytochrome P
39	122	5.6	503	1 R81464	Human derived cyto
40	122	5.6	503	1 R93170	Human cytochrome P
41	122	5.6	503	1 Y05202	Human CYP3A4 prote
42	122	5.6	512	1 R72364	Human auxillary cy
43	122	5.6	512	1 R33171	Human cytochrome P
44	121	5.6	514	1 W34539	Cytochrome P450 cc
45	120.5	5.5	493	1 R81467	Human derived cyto

## ALIGNMENTS

RESULT 1	
ID Y04128	standard; Protein: 405 AA.
AC Y04128:	
DT 11-JUN-1999 (first entry)	
DE Pseudomonas putida cytochrome P450 protein P450-cam.	
KW Bacterial; mammalian; cytochrome P450; chimeric; fusion protein;	
KW oxidase; hydrocarbon; carbon-hydrogen bond; hydroxylating;	
KW bioremediation; environmental pollutant.	
OS Pseudomonas putida.	
PN W09908812-A1.	
PD 25-FEB-1998.	
PF 17-AUG-1998; U16979.	
PR 20-AUG-1997; US-056754.	
PA (UYRP ) UNIV ROCHESTER.	
PI Jones JP, Shimoji M;	
DR WPI: 99-190131/16.	
DR N-PSDB; X19926.	
PT New P450 fusion proteins - comprising a portion of a bacterial	
PT cytochrome P450 protein and a portion of a mammalian cytochrome P450	
PT protein	
PS Disclosure: Page 12-13; 51pp; English.	
CC The present invention describes a fusion proteins comprising a portion	
CC of a bacterial cytochrome P450 protein and also a portion of a mammalian	
CC cytochrome P450 protein. The fusion protein can oxidise hydrocarbons or	
CC any compound having a carbon-hydrogen bond. The fusion protein can be	
CC used for hydroxylating a compound to be oxidised. It can also be used in	
CC the bioremediation of an environmental pollutant. Since the fusion	
CC protein is soluble, it can be subject to structural elucidation by x-ray	
CC crystallography for designing functional proteins. It can be readily	
CC expressed in soil bacteria to facilitate bioremediation. The present	
CC sequence represents Pseudomonas putida cytochrome P450 protein P450-cam	
CC from the present invention.	
SO Sequence 405 AA:	

Query Match 97.8%; Score 2133; DB 1; Length 405;  
Best Local Similarity 99.8%; Pred. No. 3.4e-209;  
Matches 404; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 10	NIAPRPVPEHLVDFDMYNSNL	SAGVQEA	AVLQESNV	VDLWTR	CNGSH	VIAT	RGQ 69
DB 1	NIAPRPVPEHLVDFDMYNSNL	SAGVQEA	AVLQESNV	VDLWTR	CNGSH	VIAT	RGQ 60
QY 70	LIREVEYDRHSSCPFP	PREAGEAYDP	IPSPMD	PPREOR	ORALANOV	GVGPPV	VDKLN 129
DB 61	LIREVEYDRHSSCPFP	PREAGEAYDP	IPSPMD	PPREOR	ORALANOV	GVGPPV	VDKLN 120
QY 130	RIQELAGSLIESLR	QOGCNFTEDYA	AEPPRIR	IFMLAGL	PREDP	PHUKYL	TDQTRPDG 189
DB 121	RIQELAGSLIESLR	QOGCNFTEDYA	AEPPRIR	IFMLAGL	PREDP	PHUKYL	TDQTRPDG 180
QY 190	SMTFAEAKALDYLI	PIITEORRQ	RPGTDAIS	IVANGOV	NGRPITS	DEAKRM	CGLLVGG 249
DB 181	SMTFAEAKALDYLI	PIITEORRQ	RPGTDAIS	IVANGOV	NGRPITS	DEAKRM	CGLLVGG 240

OY	250	LDTVNFISFSENEFLAKSPHEHOELERERERI	PACCELLRRSLVAADRILTSDEPEFG	309
Dd	241	LDTVNFISFSENEFLAKSPHEHOELERERERI	PACEELLRRFSVLADGRILTSDEPHG	300
OY	310	VOLKKGDOILLPOMLSGLDERKNACPMHVD	FROKVSHHTFGHGSHLCTGOHLAREEIV	369
Dd	301	VOLKKGDOILLPOMLSGLDERENACPMHVD	FROKVSHHTFGHGSHLCTGOHLAREEIV	360
OY	370	TLEKWLTRIPDPSIAPGAQIOHKSGIVSGVAL	PLWMDPATTKAV	414
Dd	361	TLEKWLTRIPDPSIAPGAQIOHKSGIVSGVAL	PLWMDPATTKAV	405
 RESULT 2 Y04126 Y04126 standard; Protein: 446 AA.				
ID	Y04126;			
AC	Y04126;			
DE	11-JUN-1999 (first entry)			
KM	Bacterial and mammalian chimeric cytochrome P450 protein.			
KM	Bacterial; mammalian; cytochrome P450; chimeric; fusion protein;			
KM	oxidise; hydrocarbon; carbon-hydrogen bond; hydroxylating;			
KM	bioremediation; environmental pollutant.			
OS	Synthetic.			
PN	MO9908812-A1.			
PD	25-FEB-1999			
PR	17-AUG-1998; U16979.			
PA	20-AUG-1997; US-056754.			
PI	(UYRP ) UNIV ROCHESTER.			
DR	Jones JP, Shimoji M;			
N-	WPJ: 99-190131/16.			
DR	N-PSDB: X19916.			
PT	New P450 fusion proteins - comprising a portion of a bacterial			
PT	cytochrome P450 protein and a portion of a mammalian cytochrome P450			
PT	protein			
PS	Claim 24; Page 6-8; 51pp; English.			
CC	The present sequence is a fusion proteins comprising a 'portion of a			
CC	bacterial cytochrome P450 protein and also a portion of a mammalian			
CC	cytochrome P450 protein. The fusion protein can oxidise hydrocarbons or			
CC	any compound having a carbon-hydrogen bond. The fusion protein can be			
CC	used for hydroxylating a compound to be oxidised. It can also be used in			
CC	the bioremediation of an environmental pollutant. Since the fusion			
CC	protein is soluble, it can be subject to structural elucidation by X-ray			
CC	crystallography for designing functional proteins. It can be readily			
CC	expressed in soil bacteria to facilitate bioremediation.			
SC	Sequence 446 AA:			
 Query Match 54.4%; Score 1185; DB 1; Length 446; Best Local Similarity 57.6%; Pred. No. 1.8e-112; Matches 253; Conservative 33; Mismatches 99; Indels 54; Gaps 9.				
OY	10	MLAPPHPNPELVDFDMYNPSNL	SAGVOEAMVLOESNVDDLVTWCNGSHWIATRCO	69
Dd	1	NLAAPLPNPELVDFDMYNPSNL	SAGVOEAMVLOESNVDDLVTWCNGSHWIATRQ	60
OY	70	LIREAYEDYRHFSSECPLIPREAGAAYDIPT	SMDDPEQORFALANOVGMVVYDKLEN	129
Dd	61	LIREAYEDYRHFSSECPLIPREAGAAYDIPT	SMDDPEQORFALANOVGMVVYDKLEN	120
OY	130	RIOELACSLIESLRQGOCNFTEDYAEP	PPRIIFMLLAGLPREDIPHLKYLTDOMTRPG	189
Dd	121	RIOELACSLIESLRQGOCNFTEDYAEP	PPRIIFMLLAGLPREDIPHLKYLTDOMTRPG	180
OY	190	SMTFAEAKEALDYILPIIEORRKP	G-----DAISIVANGGVNGRP--TJSSEAKKM	241
Dd	181	SMTFAEAKEALDYILPIIEORRQP	RGNNPDPDFICFLMKMEKXHNQPSFTLESLENT	240
OY	242	CGLLVGLIDTVNVLFSFSEMEFLAKSP	EHROELIERPERI-----PA	283
Dd	241	AVDLFEACTETTTSTTLARALLLLKL	HPEYTAQVOEIEIRVIGRNNSPCMQDSHMPTYDA	300
OY	284	ACEELLEFRFSLVADG--RLITSDEY	EFGVOLKKGDOILL-----POMLSG-	326

[illegible]

Query Match Similarity	16.7%	Score 364;	DB 1;	Length 396;
Best Local Similarity	28.5%	Pred. No. 7,9e-29;		
Matches 105;	Conservative 57;	Mismatches 168;	Indels 38;	Gaps 8;
QY	63	WATRGQLIREAYEDYRHFSSSECPIT-----PRAGEAYDEIPTSMDDPEQORFRA	113	
Db	43	WLVTNR-----YEDVRAVLGDGRFVRGSPMTRDEPRTPREKVMKGLISMLPSSRLNR	95	
QY	114	LANQVVGMPVYDKLENRIQELACSLIESLRPGQ-CNFTEDYAEERPPIRIFMLAGLPEE	172	
Db	96	LVMKFTARRAESLPARRAREIAHELVDQMAATGQADALVAMFAROLPVRVICELLGVP5A	155	
QY	173	DIPHLKYLTDQMTRPDGSN-----TEFAEKALYDYLPTITEOROKPGTDAISIV	223	
Db	156	D-----HDFTRKSGAFSLTAEYTAERMQZAAACATAYAMDLLDRRKKEPTDVL5AL	208	
QY	224	ANGVGNRPITSDAEKRMCGELLVGLDPTVNFLSFSMEFLAKSEPHROELIERERIPA	283	
Db	209	VQARQOQSLSRQELDLDAIGLVAGVEYSTTQIADVFYLLMTRELRLRQLLDRELIPS	268	
QY	284	ACEELLRFESL---VADGRLTSDYEFHGVLKSKDQILLPOMLSGLDERKAKACPMHVF	340	
Db	269	AVEELTRRVPLGVGTAFPRYAVEDTTLKGVTIRAGEPYLASTGAANRQAOQPPDADRIDV	328	
QY	341	SRQKVSHTTFGSHSLGCGHNLARELLIVTLKEMLTRLPDPSIA-PCAQIQKSG-IYSG	398	
Db	339	DRTFMQHLGFGGIVHHCGLAPLARELQVALLEVLRLRGLRGIPEQLRSEGMLLRG	388	
QY	399	VQALPLVM 406		
Db	389	PLELPLVM 396		

R77867  
ID R77867 standard; Protein: 408 AA.  
AC R77867;  
DT 13-NOV-1995 (first entry)  
DE S. clavuligerus ORF10 product.  
KM Clavulanic acid; clavulinate; antibiotic; beta-lactamase-inhibitor.  
OS Streptomyces clavuligerus.  
PN CA2108113-A.  
PD 09-APR-1995.  
PF 08-OCT-1993; 108113.  
PR 08-OCT-1993; CA-108113.  
PA (UYAL) UNIT ALBERTA.  
PI Aldoo KA, Jensen SE, Paradkar AS;  
DR N-PSDB: Q91580.  
PT Clavulanic acid biosynthesis enzymes and corresp. DNA - useful for  
PT biosynthesis of the antibiotic in Streptomyces hosts which do not  
PT naturally produce clavulanate  
PS Claim 32; Fig.19; 41pp; English.  
CC A 15 kb fragment S. clavuligerus NRRL 5741 genomic DNA (Q91580),  
CC extending downstream from pcbC, included 10 ORFs encoding the  
CC enzymes required for clavulinate biosynthesis. The ORF10  
CC product (R77867) showed high similarity to cytochrome P450-type  
CC enzymes from other Streptomyces spp.  
SQ Sequence 408 AA;

Query Match 15.2%; Score 332; DB 1; Length 408;  
Best local similarity 29.0%; Pred. No.1.5e-25;  
Matches 122; Conservative 59; Mismatches 185; Indels 54; Gaps 18;

QY 16 PHVPHLVDFDMYNPSNAGVOEAMVLOESNVPDLVWTCNGSH-WIATRGQLIREA 74  
DB 14 PAVPHHRCVVD---PPQLAGLRBQKASRYT-----LM---DSQVWLVSAGARAV 62  
QY 75 YEDYHFS-SECFRIP-----REAGEAYDFIPTSMPEORQFRA-----LANOV 118  
DB 63 LGDRFATVTSAPGRPMILRTSQLVRANPESASF--RMDPDQHSRLSMLTRDFLARRA 120  
QY 119 VGM-IVNVKLEKRIQELACLISLESIRPOGQCNFTEDYAEPPRIRIFMLAGLPEDDIPHL 177  
DB 121 EALRAVAVREL--LDLILGGLVKGEP--VDLVAGLTIPVPSRVITLFLFGAGDDRREPI 174  
QY 178 K-----YLTDMTRPDGSMTEFAEKALYDYLPIIEORRQKGTALISIVANGVNGRP 233  
DB 175 EDRSAVLIDRGTYPE---QVAKARDELQCYLRKELVERKENGSTDLISLVLDQVRPHLU 231  
QY 234 TSDAARRMCGILLVGLDVTNVFLSFSMEFLAKSPENRQELIERPERIPACEELLRRFS 293  
DB 232 RVEEAVPMCRLLLVAGHGTTTQASLSTLSLTDPDELGRLEDPALLPKAVEELLRFHS 291  
QY 294 LVADG--RLTSDYEFHGVOLKKGDQILLPOMLSGLDERKKACPMHNDVFSRQKVSHTTFG 351  
DB 292 IVQNGIARAIVEDVQLDVLLFRAGGVVLSAGNRNDEVLPDPVRVDYDRARRHLLAFG 351  
QY 352 HGSNCLGQHLAR---REIIVTLKEMLTFRIPDESIA-PCAOIOHKSIGVY-GVQALLPLWV 406  
DB 352 HCMHCCLOGLMARVLELELAAVLARM---PGARLAVPEEELDFRHEVSYGLGALPLVW 408

RESULT 5  
W33274  
ID W33274 standard; Protein: 587 AA.  
AC W33274;  
DT 17-FEB-1998 (first entry)  
DE S. fradiae tylosin biosynthesis gene product tyIH.  
KM Tylosin; biosynthetic gene product; production; antibiotic; tyIH.  
OS Streptomyces fradiae.  
PN US5672497-A.  
PD 30-SEP-1997.  
PF 21-DEC-1995; 575843.  
PR 12-MAY-1989; US-351350.  
PR 21-MAR-1986; US-842330.

PR 25-JUL-1986; US-890670.  
PR 24-FEB-1987; US-018237.  
PR 06-AUG-1991; US-742222.  
PR 28-JUL-1993; US-107232.  
PR 17-FEB-1994; US-198672.  
PR 21-DEC-1995; US-575843.  
PA (ELIL) LILLY & CO ELI.  
PI Cox KL, Fishman SE, Hersberger CL, Seno ET;  
DR N-PSDB: T58686.  
PT DNA encoding Streptomyces fradiae tylosin biosynthesis gene products  
PT - for increasing tylosin production in Streptomyces spp.  
PS Claim 20; Columns 21-38; 38pp; English.  
CC The present sequence is the Streptomyces fradiae tylosin  
CC biosynthetic gene product tyIH, useful to increase the production  
CC of the antibiotic tylosin in Streptomyces spp.  
SQ Sequence 587 AA;

Query Match 15.2%; Score 330.5; DB 1; Length 587;  
Best local similarity 29.3%; Pred. No.3.8e-25;  
Matches 105; Conservative 62; Mismatches 170; Indels 21; Gaps 11;

QY 63 WIATRGQLIREAYEDYR--HFSSECFIPREAGEA-YDFIPTSMPEORQFRLANOV 118  
DB 143 WLISRODHVALLADPRAVSIH-PAKLPRLSPSDGEASRSLLTDPDPHGALRGHFIPE 201  
QY 119 VGMPPVVDKLEKRIQELACLISLESIRPOG-QCNFTEDYAEPPRIRIFMLAGLPEDDIPHL 177  
DB 202 FELRRVRELRPEVSEQIVGLDLDLTARGDEADLDLAFALPMATQYICRLDIPYEDRDFV 261  
QY 178 KYLTDMTRPDGSMTEFAEKALYDYLPIIEORRQKGTALISIVANGVNGRPITS 235  
DB 262 QERTDQATPPAAGEALELLELRDYLRLISGKTGREGSDGLSMVA--OARGGGLSH 319  
QY 236 DEARRMCGILLVGLDVTNVFLSFSMEFLAKSPENRQELIERPERIPACEELLRRFSIV 295  
DB 320 ADVLNAVALLAAGHETTASMTGSLVYLQHPTRAMRELTVPGLPGAVDBLLAYLS-I 378  
QY 296 ADG--RLTSDYEFHGVOLKKGDQILLPOMLSGLDERKKACPMHNDVFSRQKVSHTTFG 353  
DB 379 ADGLRSATADLEIDGHIRAGDGLVFLLAANRDEAVSEPEADPIHRSARRHAAFGV 438  
QY 354 SHLCIGQHLARREIIVTLKEMLTFRIPDESIA-----GAOIOHKSIGVY-GVQALLPLWV 406  
DB 439 PHQCLGQNLARMELEVALGAVLERLP--ALRPPTYAGRLKSDSA-VFGVTELPVW 493

RESULT 6  
R11349  
ID R11349 standard; Protein: 406 AA.  
AC R11349;  
DT 05-JUN-1991 (first entry)  
DE Cytochrome enzyme P450SUL.  
KM Cytochrome P450; P450SUL; P450SU2; herbicide resistance.  
OS Streptomyces griseolus.  
PN W09103561-A.  
PD 21-MAR-1991.  
PF 27-AUG-1990; 004785.  
PR 11-SEP-1989; US-405605.  
PR 12-JAN-1990; US-464499.  
PR 23-AUG-1990; US-569781.  
PA (DUPO) DU PONT DE NEMOURS CO.  
PI Dean C, Harder PA, Leto KJ, Lichtner FT, Odell JT;  
PI O'Keefe DP, Omer CA, Komesser JA;  
DR WPI: 91-102077/14.  
DR N-PSDB: Q11126.  
PT DNA encoding cytochrome P450 enzymes - and electron donating  
PT iron-sulphur proteins, used to confer herbicide resistance to  
PT plants and microorganisms  
PS Claim 13; page 151; 224pp; English.  
CC This cytochrome P450 enzyme, P450SUL is expressed alongside the iron  
CC sulphur protein Fes-B, by a DNA sequence contained in a recombinant

CC plasmid. Host *Streptomyces* species are transformed with the plasmid  
CC and are used to coat a plant seed to transform the plant. The res-  
CC ultant transformants are resistant to herbicides.

CC See also Q11127.  
SQ Sequence 406 AA;

**SQ Sequence 406 AA;**

Query Match	15.1%	Score 330;	DB 1;	Length 406;
Best Local Similarly	26.38%	Pred. No. 2.4e-25;		
Matches 104; Conservative	74;	Mismatches 193;	Indels 24;	Gaps 11;

QY	31	PSNLSAGVO--EANAIVLOESVPLDWTMRCGGH-WIATRQOLI REAYEYR-----	79
Db	17	PSNNSCPQLDEGTAQQLRDTGPIRLHRTYLDVGRQAMVTKHKAARKLLGDRUSSNRTDD	76
QY	80	HFSSECPITP--REAGEAYDEIPTSMDPPEEOROFALANQVCAVPVVDKLENNIOELACS	137
Db	77	NFPATSPFEAVRESQAF-----IGDRPELGTRRRMTISFTYKRTKGMARPEVEEYVHG	132
QY	138	LIESLRPOG--OCNFTEDYAEFPPIRIEMLAGLEEDIPLIKYLTDMOTRPGSGMTPEA	166
Db	133	FLDEMLAGPADIIVSOPALPVPNMVICRLIGVYADHEFPQDASKRVLQSTDAQSALTA	192
QY	197	KEALYDYLIPIIEQRQKRGTDAL-STYANQVNGRPTSDSEAKRMCGLLVGLDTPVN	255
Db	193	RNDLAGVYDGLITPQYEPERGAGVLGALVAADOLANGE-IDRELLSTAMLLIAGHETTAS	251
QY	256	FLFSFMEFLAKSPHROELIERPERIPACCELLRRFSL--VADGRILTSDYEFHGVLK	313
Db	252	MTSLISVITLLDHPROXYALNRADRSRLVCGAVBELLRYLAIADIAGGRATADIIEVGHILIR	311
QY	314	KGDQILPQMLSGLDERKKNACPMHVDFSROKVSHTTTFGSHSLCGLGHLARREIIVTKE	373
Db	312	AGEGIVVNSIANDRGVYEDPDALDIHRSARHHLAGFGVQCGLONLARLELEVITLNA	371
QY	374	WLTRIPDESIA-PCGAQLOHKSG-IVSQVQALPLVM	406
Db	372	LMDRVPITRLAVPVEQLVLRGTTIOGNEPLVTM	406

RESULT 7  
W54389  
ID W54389 standard; Protein; 411 AA.  
AC W54389;  
DT 18-AUG-1998 (first entry)  
DE Actinomadura habisica polyketide synthase protein 10.  
KW Multienzyme; infection; fungi; yeast; gram-positive bacteria;  
KM dihydrotolbenzo(a)naphthacenequinone aglycon; antibiotic; pradimicin.  
OS Actinomadura habisica.  
PN W09811230-A1.  
PD 13-MAR-1998.  
PE 13-SEP-1996; U14791.  
PR 13-SEP-1996; WO-U14791.  
PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
PI Dairi T, Oki T;  
DR WPI; 98-207391/18.  
DR N-PSDB; V26609.  
PR P-PSDB; W54388, W54389, W54390.  
PT Actinomadura polyketide synthase genes - useful for preparation of  
PT pradimicin.  
PS Disclosure; Page 55-56; 71pp; English.  
CC The Actinomadura habisica polyketide synthase proteins W54380-W84390 form  
CC a multienzyme complex. The enzyme is used for the biosynthesis of a  
CC dihydrotolbenzo(a)naphthacenequinone aglycon preferably a pradimicin which  
CC is an antibiotic useful against systemic fungal infections caused by  
CC Candida albicans, Aspergillus fumigatus and Cryptococcus neoformans.  
CC It is also active against a wide variety of fungi and yeasts, some  
CC Gram-positive bacteria and viruses.  
SQ Sequence 411 AA;

Query Match	15.18;	Score 329;	DB 1;	Length 411;
Best Local Similarity	26.48;	Pred. No. 3.1e-25;		

Matches	114:	Conservative	61:	Mismatches	199:	Indels	58:	Gaps	16:
QY	4	TTISANIANAPLPHPVPEHLVDFDMYNP----	SNLSAGVQEAAMVLOESNVDLWMTKCN	59					
	:	:	:	:	:	:	:	:	:
Db	9	TVDPRPDVTPAFPEPRD-----	DFQPPCCHARKLRASDPYAKVLP-----	T	50				
QY	60	GGH-WIATRGQLIRAYEDYRHFSSEC----	BEIPEAGEAVDELPTS-----	MDPE	107				
	:	:	:	:	:	:	:	:	:
Db	51	GDHAWVTRIVADRVFVSD-RRFSKEAVTRGAPR-----	LIPMORGSLVIMDPE	102					
QY	108	QOQFRAIANQVGMPEVDKLENRIQELACLSTLESIPQG--	OCNFTEDYAEPPPIRIFML	166					
	:	:	:	:	:	:	:	:	:
Db	103	HTRMKRIYSRAFTARRVGEGRAMHAYNDLTSGFVDEKVEHGPRADLIHALPLPVTVICEM	162						
QY	167	AGLPREDIPHLKYLTDQOTRDRGDSMTFAEANE-----	ALDYLIPIIEQRQRPGDAI	220					
	:	:	:	:	:	:	:	:	:
Db	163	LGVPEDPRPDQWMDTRNL-TIGAPALAAQADEIKAAVGRLYGAYIELLDATTAAPADLL	221						
QY	221	SIYANGVQNGRPITISDEKRMKCGLLVGGDLTVVNFLSFSMEFLAKSPENHQELIERPER	280						
	:	:	:	:	:	:	:	:	:
Db	222	SLSLSAHND--GLSBEELLTFCGMLTLAAGYHTTAATHSVYHLLREPSRVARLRREDPSG	280						
QY	281	IPACEELLRRPSLVADG----	RILTSYEFHGVOLKKGOIILLPQMLSGIDERRNACPM	336					
	:	:	:	:	:	:	:	:	:
Db	281	IPAAAEELL-RYGOIGGGAGAIRAVEDEVGGGLVYRGEAVNIPLFNMAANDPEVFADPE	339						
QY	337	HYDESROKVSHTTTCGSHGLCGHILARREITVTLKEMLTRIPESIA-PCAQIOHKSQI	395						
	:	:	:	:	:	:	:	:	:
Db	340	ELDDGRTDNPHALIGHGHCYGLCAPLARLEQLVYLETIVERTPALRLAIDADITWRRGL	399						
QY	396	V-SGVQALPLTW	406						
	:	:	:	:	:	:	:	:	:
Db	400	AFARPDALPIAM	411						

RESULT 8

ID W11585 standard; Protein: 398 AA.  
AC W11585;  
DT 02-APR-1997 (first entry)  
DE Streptomyces pristinaespiralis snbF gene product.  
KW Streptogramin B; antibiolic; biosynthesis; pristinamycin;  
KW virginiamycin; piperolic acid; cyclodeamination; papa; snba; snbF;  
KW p1pA; 3-hydroxypicolinic acid; hydroxylation.  
OS Streptomyces pristinaespiralis.  
PN M09601901-A1.  
PD 25-JAN-1996.  
PF 04-JUL-1995; F00889.  
PR 08-JUL-1994; FR-008478.  
PA (RHON ) RHONE-POULENC RORER SA.  
PI Barriere JC, Blanc V, Blanche F, Crouzet J;  
PI Debussche L, Paris JM, Thibaut D, Bamas-Jacques N;  
PI Dulruc-Rosset G, Famechon A;  
DR WPI; 96-097631/10.  
DR N-PSDB: T58555.  
PT New streptogramin B derives, useful as antibiotics - produced by new  
PT mutants of Streptomyces having altered genes for streptogramin B  
PT biosynthesis  
PT  
PS Example 1; Page 113-114; 146pp; French.  
PS The papa gene of S.pristinaespiralis is involved in the biosynthesis  
CC of 4-dimethylamino-L-phenylalanine (DMPPA), a precursor for  
CC pristinamycin IA. Upstream of the papa gene, on the complementary  
CC strand, is the snba gene coding for 3-hydroxypicolinic acid-AMP ligase  
CC The region between these two genes was sequenced and two open reading  
CC frames were identified. The first (p1pA) decodes to an amino acid  
CC sequence with homology to ornithine cyclodeaminase from Agrobacterium  
CC tumefaciens. The p1pA gene product is likely to catalyze the cyclo-  
CC demination of lysine, leading to production of piperolic acid.  
CC Mutations in the p1pA gene were shown to affect piperolic acid  
CC synthesis but not the synthesis of 3-hydroxypicolinic acid. The second  
CC open reading frame (snbF) could be decoded to give a product with  
CC homology to hydroxylases of the cytochrome P450 type.  
CC Disruption of the p1pA and snbF genes can be used to produce

CC strains of *S. pristinaespiralis* which are unable to produce the  
CC antibiotic pristinaemycin I but which may be able to produce new,  
CC modified forms of it.  
50 Sequence 398 AA;

Query Match	14.98	Score 324.5	DB 1	Length 398
Best Local Similarly	29.08	Pred. No. 8.5e-25		
Matches 106; Conservative	59	Mismatches 159	Indels 41	Gaps 12

```

QY 74 AYEDYRH-----SSSECPPIREGEAEYDIPISMDPEERORORALANQVGM 121
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 36 AFHFRIADVLTVASDGVYSSQISRLRPGSQALSEDILSIDPMMHRLRLRSQAFTF 95
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 122 PYVOKLEMRIOELACLSTLIESLRPGQC--NFTEDVAFEPFPIRIFMLAGLPEEDIPHLKYL 180
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 96 RTVADLERIYETLAGQLLDVA--DGDIFDLVADFAIYFLPIYIAELLGVPADPRTLFRSM 153
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 181 TDQMT-----PDGSMTEFAEKALYDLPIIEORORPGDAISIVA 224
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 154 SDRMLQMVADPADMQFGDADDEDYQFLVKEPMAMHAYLHDVHTDRARAPANDLSALV 213
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 225 NGQVNGRPITSDEKKNRKCGLLVGGLDVTYVNFISFSMEFLAKSPENHOELTERPER--1P 282
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 214 AAREGERLTDEQIYERFEGALLMAGHSTSMLGNTVCLDKDHP--FAEFAAARADSLIP 271
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 283 AACCELLR-RPSLVAADRIITSDVEFHYGOLKGDQDITLQPMIS--GLDERKNACPMHVD 340
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 272 ALITEVLRKRPPIYVMARVYTKDITVLACTIIPAC--KNVVSLSLSANHDQVFTDPDHLDL 330
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 341 SRQVSHTEFGHSGHSLGCHLARREIIVTLKEWLTIRIPDSIAPGAQO--KRSGIYSGV 399
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 331 AREG-RQIARFGCHGYCGAPLARLEGRIALDEALFDBFDPFSPDGAKLKYHHNDGLF--GV 388
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 400 QALFL 404
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 389 KNEPL 393
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

RESULT 9  
 ID R11350 standard; Protein; 403 AA.  
 AC R11350;  
 DT 05-JUN-1991 (first entry)  
 DE Cytochrome enzyme P450SU2.  
 KW Cytochrome P450; P450SU1; P450SU2; herbicide resistance.  
 OS Streptomyces griseolus.  
 PN M09103561-A.  
 PD 21-MAR-1991.  
 PF 27-AUG-1990; U04785.  
 PR 11-SEP-1989; U5-405605.  
 PR 12-JAN-1990; U5-464499.  
 PR 23-AUG-1990; U5-569781.  
 PA (DUP0 ) DU PONT DE MEMOURS CO.  
 PI Dean C, Harder PA, Leto KJ, Lichtner FT, Odell JT;  
 PI O'Keefe DP, Omer CA, Romesser JA;  
 DR WPI; 91-102077/14.  
 DR N-PDSB; Q11127.  
 PT DNA encoding cytochrome P450 enzymes - and electron donating  
 PT iron-sulphur proteins, used to confer herbicide resistance to  
 PT plants and microorganisms  
 PT Clalm 15; page 158; 224pp; English.  
 PS This cytochrome P450 enzyme, P450SU2 is expressed alongside the iron  
 CC sulphur protein FeS-, by a DNA sequence contained in a recombinant  
 CC plasmid. Most Streptomyces species are transformed with the plasmid  
 CC and are used to coat a plant seed to transform the plant. The res-  
 CC ultant transformants are resistant to herbicides.  
 CC See also Q11126.  
 SQ Sequence 403 AA;

Query Match	14.8%;	Score 322;	DB 1;	Length 403;
Best Local Similarity	28.4%;	Pred. NO.	1.6e-24;	

Matches 103; Conservative 64; Mismatches 166; Indels 30; Gaps 13;

QY	63	MIATPGCOLIREAYEDYHRESEC-----PIPIRAGAGAYDPIR--SMDEPEQOROPALA	115
Db	52	MLVTRHODVRAVLGPPR-FSADAHRTGEPPLTTACGREGICTNPFLRMDPEHARLRML	110
QY	116	NOVGMPPVADLENNRIEOLACSLIESLRP-OGOCNFEDYEAPEPPIPIFMIALGLPEEDI	174
Db	111	TADFIYKKEVEMAREPVQGLADOLVDRMHTGTSADLTETAPLRPLSLVICLLGVPREDH	170
QY	175	PHL----KYLTIDQMRPDGSMTFPAEAKALYDYLIIPIEOROKPGTDAIS-IVANQVN	229
Db	171	AFFOERSVLLTLTRSTPE---EVRAADELLEYELARLARTRKEPRDDAIIISRIVARGELD	227
QY	230	GRPTISDIAKRCGGLLYGVGLDYVNFLSFMEFLASPEHROELIRPERIRIACCELL	289
Db	228	DTQIAT-----MGRLLVAGGELDDDTQIATMGLLNRPOLARLRAPALVAGCAVEELL	282
QY	290	RRESFLVADG--RIITSDYEFHGYOLTKGDOILLPQMLSGDERKNACP--MHNDEFOSROKV	345
Db	283	RYLTIIVHNGVRIRATEDVYLIGCRITLAGEGYLC--MISSANRADEYVPGGDDLDVARDAR	340
QY	346	SHHTFGHSGHLLCLOHRLARREIYTLKEMILTRIPDFESA--PGAQIOHKSGI--VSGOALP	403
Db	341	RHVAFGFCVHQCLOGLAPLARVELQIAIFELRLRDLRLAVRHEEIPPRGMALYGVHSLP	400
QY	404	LVM 406	
Db	401	IAM 403	

RESULT 10

ID R38309 standard; Protein; 412 AA.

AC R38309;

DT 04-DEC-1993 (first entry)

DE Sequence of the P450-soy protein.

KM P450soy; soyG gene; oxidation; haem protein.

OS Streptomyces griseus ATCC 13273.

PN W09J12236-A.

PD 24-JUN-1993.

PE 16-DEC-1992; U10885.

PR 16-DEC-1991; US-807001.

PA (DUPO ) DU PONT DE NEMOURS & CO E. I.

PI Omer CA, Sarislani FS, Trower MK;

DR WPI: 93-214178/26.

DR N-PSDB: 045569.

PT Constitutive expression of P450 SOY (SOYG) and ferredoxin soy (soyB) in Streptomyces - used for oxidn. of organic chemicals

PS Example: Figure 2; 45pp; English.

CC Cytochrome P450soy was purified from *S. griseus* ATCC 13273. Two similar forms of P450soy were isolated. P450soy-delta, is derived from P450soy by *in vitro* proteolysis during isolation. One of the tryptic peptide fragments of cytochrome P450soy and of the P450soy-delta protein were subjected to automated degradation. The NH2 terminal sequences are given in R38306 and R38307. A mixture of CC oligos that consist of possible DNA sequences that could encode the AAs FGVHQL of the tryptic peptide was made. It consists of the CC sequences in Q43290-23. The oligo mixture was end-labeled and used to probe the EMBL4 library of *S. griseus* DNA. Hybridising plaques were isolated and a 4.8kb SacI DNA fragment was isolated from one CC clone that hybridised to the oligo probe mixture. As segment of the CC 4.8kb fragment was sequenced and found to contain an ORF. Within CC this ORF was a section that corresp. exactly to the AA sequence determined from the cytochrome P450soy tryptic peptide (see Q45569, CC R38309). The gene encoding the P450soy protein was called SOYG. Five CC nucleotides downstream from the stop codon for SOYG another ORF was identified. This ORF encodes an apparent ferredoxin-like protein.

CC The gene was designated soyB and the protein ferredoxin-soy.

SO Sequence 412 AA;

Query Match	14.58;	Score 316;	DB 1;	Length 412;
-------------	--------	------------	-------	-------------

Best Local Similarity 27.2%; Pred. No. 6.6e-24;  
Matches 84; Conservative 56; Mismatches 163; Indels 6; Gaps 5;

QY	103	MDPEQRQFRLANOVVGMVYDKLENRIGOLASGLIESLRPOS-QCNFTEDVAEPPIR	161
	104	103	103
Db	105	VDDPEHNTQRMMILPTTSVYKIGALRRPRLTQTYVDRLIDMMRQOPRAPELVSAFRLPSPM	164
	106	105	105
QY	162	IFMLLAGLPEEDIDPHLKLYLTDQMTRPDGSMTFAEAKALYDYLPIITEOROKPGTDAIS	221
	163	162	162
Db	165	VICALGVYPVADHNAFFEEERSQRLRLCGADVADNARABELEUYGALIDRRKRAEPGSDLLD	224
	166	165	165
QY	222	IVANGOVVGRPTSDGEAKRMCGLLVYSGLDLPVVNFLEFSMBFLAKSPENHOELIERERI	281
	223	222	222
Db	225	ELIRHDHDEGVYDREQLVAFVILLIGHETIYAMISLIGFTLLSHPEQLAALRAGST	284
	226	225	225
QY	282	PAGECELLRFSLVADG--RLTSDYEPFHGVOLKKGOULLPOMLSGIDERRKNACPMHVD	339
	283	282	282
Db	285	AVVVVEELL-RELSTIAEGIORLATATEDMEVDGATIRKGGVVFESTLIRNDADVPRATLLD	343
	286	285	285
QY	340	FSRQKVSHTTTFGHGSHLCLGOHLARRETIYALKWMLRIRIDFESIA-NGAOTIOHKSG-IYS	397
	341	340	340
Db	344	MDRRARHHILAFGFGVHOCISGONLARLELDIMKRTLFRFLRGLRAYVAMETIRKRPGETTQ	403
	345	344	344
QY	398	GVQALPLVM 406	
	399	398	398
Db	404	GILDLPLVAM 412	
	405	404	404

RESULT	11
ID	R51368
AC	R51368 standard; Protein; 410 AA.
DT	24-NOV-1994 (first entry)
DE	Protein containing Cytochrome P450 SCA-2 activity.
KM	Cytochrome P450 SCA-2; Streptomycetes carboxiphilus;
OS	treatment of hyperlipidaemia; drug preparation.
FT	Streptomycetes carboxiphilus.
FT	Key
FT	Location/Qualifiers
FT	1. 410
PN	/label= cytochrome P-450 containing activity
PD	J06070780-A.
PD	15-MAR-1994.
PF	28-AUG-1992; 229969.
PR	28-AUG-1992; JP-229969.
PA	(SANTY ) SANKYO CO LTD.
DR	WPI: 94-128679/16.
DR	N-PSDB: Q61452.
PT	Cytochrome P450 SCA-2 gene - from Streptomycetes carboxiphilus,
PT	useful in treatment of hyperlipidaemia
PS	Claim 1; Page 12-14; 18pp; Japanese.
CC	R51368 shows a protein having cytochrome P-450 activity. P-450
CC	SCA-2 can be prepared commercially for use in the preparation of a
CC	drug for the treatment of hyperlipidaemia.
CC	Sequence 410 AA;

Query Match	14.4%;	Score 313.5;	DB 1;	Length 410;
Best Local Similarity	25.6%;	Pred. No. 1.2e-23;		
Matches 92;	Conservative 63;	Mismatches 185;	Indels 19;	Gaps 7;

[illegible]

Db	232	EIDRELLISTALLNIVACHETTAA <sup>1</sup> MTSL <sup>2</sup> SVTTLLLEN <sup>3</sup> DD <sup>4</sup> NAL <sup>5</sup> RAD <sup>6</sup> SL <sup>7</sup> YR <sup>8</sup> GA <sup>9</sup> VEELL <sup>10</sup> RV	291
Qy	292	FSL--VADGRILTS <sup>1</sup> DYEF <sup>2</sup> HVOL <sup>3</sup> KK <sup>4</sup> DO <sup>5</sup> ILL <sup>6</sup> POM <sup>7</sup> LS <sup>8</sup> GDER <sup>9</sup> KNAC <sup>10</sup> PM <sup>11</sup> IV <sup>12</sup> DS <sup>13</sup> ROK <sup>14</sup> YSH <sup>15</sup> TT	349
Db	292	LAIADIAGRIAT <sup>1</sup> ADI <sup>2</sup> EID <sup>3</sup> Q <sup>4</sup> LL <sup>5</sup> RAGE <sup>6</sup> VI <sup>7</sup> W <sup>8</sup> NSI <sup>9</sup> AR <sup>10</sup> DS <sup>11</sup> SVE <sup>12</sup> FP <sup>13</sup> RD <sup>14</sup> VR <sup>15</sup> NR <sup>16</sup> SR <sup>17</sup> AN <sup>18</sup> HL <sup>19</sup> LS	351
Qy	350	FGHSHCLSG <sup>1</sup> CHLAR <sup>2</sup> REIT <sup>3</sup> VL <sup>4</sup> KML <sup>5</sup> IRI <sup>6</sup> PR <sup>7</sup> FS <sup>8</sup> IA--FGA <sup>9</sup> DIQ <sup>10</sup> NSG--IV <sup>11</sup> SGV <sup>12</sup> AL <sup>13</sup> PL <sup>14</sup> YM	406
Db	352	FGYGVH <sup>1</sup> Q <sup>2</sup> Q <sup>3</sup> GN <sup>4</sup> LAR <sup>5</sup> LE <sup>6</sup> VL <sup>7</sup> IV <sup>8</sup> FL <sup>9</sup> PR <sup>10</sup> IR <sup>11</sup> PL <sup>12</sup> RL <sup>13</sup> AV <sup>14</sup> RE <sup>15</sup> DI <sup>16</sup> TL <sup>17</sup> PR <sup>18</sup> GT <sup>19</sup> IG <sup>20</sup> VE <sup>21</sup> LL <sup>22</sup> PT <sup>23</sup> W	410

RESULT 12

ID R14724 standard; Protein; 404 AA.

AC R14724;

DT 28-JAN-1992 (first entry)

DE 6-hydroxylase encoded by Eryf gene.

KW C-6 hydroxylation; erythromycin; 6-deoxyerythromycin; antibiotics;

KW saccharopolyspora; cytochrome P450 monooxygenase; ss.

OS Saccharopolyspora erythraea.

PN MO9116334-A.

PD 31-OCT-1991.

PF 16-APR-1991; U02600.

PR 18-APR-1990; U5-510483.

PA (ABBO ) ABBOTT LABORATORIES.

PI Weber JM;

DR WPI: 91-339744/46.

DR N-PSDB; Q14548.

PT New 6-deoxyerythromycin derivs. - are antibiotics with increased

PT acid stability, produced by cultivation of saccharopolyspora.

PS Disclosure; Fig 3; 56pp; English.

CC The eryf gene encodes the 6-hydroxylase component of the cytochrome

CC P450 monooxygenase system responsible for the hydroxylation of 6-de

CC oxyerythronolide B to erythronolide B. Interruption of this step

CC results in the formation of deoxyerythromycin A and new derivatives

CC useful as antibiotic which have better stability against acids that

CC the corresponding erythromycins. Interruption of the reaction can

CC be effected by an insertion into the eryf gene of a plasmid, gene

CC replacement or chemical or light-induced mutagenesis.

CC See also Q14549.

Q0 Sequence 404 AA;

Query Match	13.6%	Score 296.5	DB 1	Length 404
Best Local Similarity	24.2%	Pred. No. 6.3e-22		
Matches 90	Conservative 74	Mismatches 173	Indels 35	Gaps 10

```

QY 63 MIATRGOLIRATVADYDNHFS-----ECPLPRAG---EAYDPTISM---DPREQ 108
D 40 WLVTGYDEAKAALSDBLRSSDPKKRYGVEYEF-PAYLGEPEDEVNRFAJNMGTSDPPTH 98
QY 109 ROFRALANOVYGMVBYVKLEIRIOELACSLIESLRPGOCNFTEVDYEPPIRPMILAG 168
D 99 TRLKRLYSQETTVARVLEAMRPRVQIATIELLDEYDGSQVYDIYRRFHHPLPIYICELG 158
QY 169 LPEDDIPHLKYLTD-----QMTRPDGSMTFAEAKALDYLIPIEOROKPGTDAI 220
D 159 VDE-----KRGEGFRMSSLLYMDPERADRGQARREYVNLIDYVERRRTRPEDDLL 212
QY 221 SIVANGO-VNGRPITSDAEAKRMCGLLLVGSLDVTYVNFISFSMEFLAKSPENHOELIERP 278
D 213 SALIRVQDDDDGR-TSADLELSSIALVLLANGFEASVSLIGITYLLTLHPQALAVRRDP 271
QY 279 ERIPACEBELLRRSL-VADGRILTSYERHGVQLKGGDQILPQMSIGLDERKNACPMH 337
D 272 SALPNAVEELIKRYALAEQETTTTFEAEDLEIGVAIPQYSTVLVANGAANDPKQFPDPHR 331
QY 338 VDFSRKVSHTTEGSHLIGOLARREIYVLLKEVLTPIRPFSLAPGQ--IQKSGI 395
D 332 SDVYTRDTRGHSFGQIHFCMGRPLAKLEGVALRALGFRFPALSLGIDDDVYVRRSVL 391
QY 396 VSGVALPLVMD 407

```



Db 392 LRGIHLPVRLD 403

## RESULT 13

R47521 standard; protein: 398 AA.  
ID R47521:  
AC R47521:  
DT 22-JUN-1994 (first entry)  
DE Vitamin D hydroxylase.  
KW Substitution; transformation; 1-alpha, 25-dihydroxyvitamin D.  
OS Actinomycetes.  
PN W09400576-A.  
PD 06-JAN-1994.  
PE 24-JUN-1993; J00856.  
PR 25-JUN-1992; JP-167644.  
PI (TAIS ) TAISHO PHARM CO LTD.  
PI Adachi T, Bepu T, Hanada K, Horinouchi S, Kawachi H, Sasaki J:  
DR WPI: 94-026212/03.  
DR N-PSDB: Q53997.  
PI DNA encoding vitamin-D hydroxylase - obtd. from actinomycetes, used to produce vitamin-D3 and related cpds.  
PS Disclosure: Fig 6-7; 53pp; Japanese.  
CC The sequence is that of a protein from actinomycetes which is the enzyme which substitutes the H at position 25 of vitamin D cpds.  
CC with a hydroxy group (vitamin D hydroxylase). Microorganisms transformed with the gene encoding this enzyme may be used to manufacture vitamin D derivatives.  
SQ Sequence 398 AA:

Query Match 8.7%; Score 190; DB 1; Length 398;

Best Local Similarity 23.6%; Pred. No. 4.4e-11;

Matches 95; Conservative 63; Mismatches 197; Indels 48; Gaps 16;

OY 31 PSNLTAGVVO--EAMAVLOESNVPDLVWTRCNGSH-WI-----ATRGOLIREAYEDYRHPS 82  
DB 17 PSNRNAPRYFSGYAOQLRPPGRLGTRYDGRQAMVYVTHNAR-KILASFESADNR 75  
OY 83 SECPPIPREAGAYDFTPT--SMDPEORQFRALANQVGMVVDKLENRIQELACSLIE 140  
DB 76 TGFPLTAGGRDMIGNPFLFRDDPEHARLRMTISHTYARIKITISEFEVEVYLFLR 135  
OY 141 SLRPGQCNETDYAEPPI----RIFMLAGLPREDIPLH-----KYLTQMTPRDGSMT 192  
DB 136 RMTISGP--TADLVSOFLIGAHVWICRLGVPYEDNAPFQERSVLLTLSTPE---E 189  
OY 193 FAEAREALYDYLPIIEORROKPGTDATS-IVANGOVNCRPTSDPAKRMCGLLLVGSD 251  
DB 190 VAAADDELLQYLAARLAKRKRRPDATISRLVARGELREEDISDM-----LLIAGHA 244  
OY 252 TVVNFLSFSMEFLAKSPENROELIERPERIPACCELLRFSVLVADGRILTSDFEHNQV 311  
DB 245 TDASTLSVLAABRSEGYALRADRSLVPGAVEDPL-RYVALND--IATEDVPIRGHT 301  
OY 312 LKKGQOILLPOMLSGLDERKNACPMHVDPSROKVSHTTFGHSILCLGQHAR----REI 367  
DB 302 IAAQWTOIAMGSSANDHREYEDPDALPARVBDLVHLAFRTKREOCIGNLARADQLRED 361  
OY 368 IYTLKEMLRIPDFSIAPGAQIOHSGI-----VSGVALPLTW 406  
DB 362 LMTARN-----DPTLRILAVPOSSARLEGTTIGVNEPLVTW 398

## RESULT 14

ID W36128 standard; protein: 422 AA.  
AC W36128:  
DT 21-MAY-1998 (first entry)  
DE Daunomycin C-14 hydroxylase.  
KW Daunomycin C-14 hydroxylase; doxa gene; doxorubicin; daunomycin; 13-dihydrocarcinomycin; carminomycin; anthracycline;

KM anticancer; cytostatic; cancer; therapy.  
OS Streptomyces sp. strain C5.  
PN W09744439-A2.  
PD 27-NOV-1997.  
PE 22-MAY-1997; U08690.  
PR 24-MAY-1996; US-653650.  
PA (OHIS ) UNIV OHIO STATE RES FOUND.  
PI Desanti CL, Dickens ML, Strohl WA:  
DR WPI: 98-018495/02.  
DR N-PSDB: V01447, V01451-53.  
PT Producing doxorubicin from daunomycin C14 hydroxylase  
PI - also hydroxylation and oxidation of other anthracycline(s) with the same enzyme  
PS Claim 9; Page 31-33; 59pp; English.

CC This protein comprises the daunomycin C-14 hydroxylase (DC14H)  
CC of Streptomyces sp. strain C5. DC14H is a p450-like enzyme useful  
CC for the production of doxorubicin from daunomycin, and for the  
CC hydroxylation and oxidation of other anthracyclines. Its amino  
CC acid sequence was deduced from the doxa gene (see V01447)  
CC identified in the daunomycin biosynthesis gene cluster from C5 (see  
CC V01453). Plasmid vectors (see V01451-52) have been constructed  
CC for use in the recombinant production of DC14H in transformed host  
CC cells. Claimed methods for production of doxorubicin include: the  
CC use of a host microorganism transformed with a plasmid which  
CC contains the doxa gene; incubating a culture of Streptomyces sp.  
CC strain C5 with daunomycin; and incubating DC14H with daunomycin.  
CC DC14H can also be used in claimed methods of producing  
CC 13-dihydrocarcinomycin and carminomycin from 13-deoxydaunomycin,  
CC and of producing 13-dihydrodaunomycin and daunomycin from  
CC 13-deoxydaunomycin. The methods preclude the use of halogens as  
CC required in the current chemical methods.  
SQ Sequence 422 AA:

Query Match 7.3%; Score 160; DB 1; Length 422;

Best Local Similarity 23.6%; Pred. No. 5.6e-08;

Matches 96; Conservative 51; Mismatches 205; Indels 54; Gaps 14;

OY 39 QEAMAVLOESNVPDLVWTRCNGSHWIATRGOLIREAYEDYRHPSF--CPPIPREAGAY 96  
DB 31 REAGPVV--EVNAP-----AGGPRAWITDDALAREVLADPRFKGDLAPFTAMRGVDDGL 83  
OY 97 D-----FITSMDPRQROFRALANQVGMVVDKLENRIQELACSLIE-----R 143  
DB 84 DIPVRELPRFTLLIADGEBHRRLRNHAPENRPRRAEFTDRIATLADRLTELADSSDR 143  
OY 144 POGQCNETDYAEPPIRIFMLAGLP-----EEDIHLYLV-----TDQWTRPD 188  
DB 144 SEBPALLEGFRVHNPFLVITCELVPTDPRAMARAVALKALGIGSQSAGDSTDA 203  
OY 189 GSWTFAEAREALYDYLPIIEORROKPGTDATISIVANGOVNCRPTSDPAKRMCGLLLVG 248  
DB 204 GVPPTSALESF--LLEAVHNAARKQRTPTMTVLYERAQAEFGVSDDQLVVMITGLIFA 261  
OY 249 GUDTVVNFLSFSMEFLAKSPENROELIERPERIPACCELLR-----FSLVADGRILTS 303  
DB 262 GHDYTVNFI--LLEVLAGRLAADADGDAISREVEBALRHNHPVPYSL--RFFAAT 315  
OY 304 DYEPNGVQLOKDOILLPOMLSGLDERKNACPMHVDPSROKVSHTTFGHSILCLGQHAR 363  
DB 316 EYVINGVRLPRAPVLYVDEGTNTDGRNHDAFHAFPRPSRRRLTGGCGPYTCIGEOIA 375  
OY 364 RREIIVTLKEMLRIPDFSIA--PGAQIO--HKSGLVSGVALPLTW 406  
DB 376 QLESRTMIGVLSRFPQARLAVPYEELRWCRGAGQARTLTDLP-VW 420

## RESULT 15

ID W36132 standard; protein: 474 AA.  
AC W36132:  
DT 21-MAY-1998 (first entry)  
DE N-terminal modified daunomycin C-14 hydroxylase.

KW Daunomycin C-14 hydroxylase; doxa gene; doxorubicin; daunomycin;  
KW 13-dihydrodaunomycin; carminomycin; anthracycline;  
OS anticancer; cytostatic; cancer; therapy; plasmid pANT199.  
OS Chimeric - Streptomyces sp. strain C5.  
OS Chimeric - synthetic.  
FH Key Location/Qualifiers  
FT Cleavage\_site 31  
FT Protein /note= "enterokinase cleavage site"  
FT 53.474  
FT /note= "native daunomycin C-14 hydroxylase"  
PN W09744439-A2.  
PD 27-NOV-1997.  
PE 22-MAY-1997; U08690.  
PR 24-MAY-1996; US-653650.  
PA (OHIS ) UNIV OHIO STATE RES FOUND.  
PI Desanti CL, Dickens ML, Strohl WA;  
DR WPI: 98-018495/02.  
DR N-PSDB: V01452.  
PT Producing doxorubicin from daunomycin by daunomycin C14 hydroxylase  
PT - also hydroxylation and oxidation of other anthracycline(s) with  
PT the same enzyme  
PS Disclosure; Fig 9; 59pp; English.  
CC This protein comprises an N-terminal modified enzyme derived from  
CC the daunomycin C-14 hydroxylase (see W36128) of Streptomyces sp.  
CC strain C5. It is encoded by a gene construct (see V01452) in  
CC plasmid pANT199. In this plasmid, the doxa gene (see V01447) is  
CC translationally fused with a leader sequence encoding 6  
CC histidine residues so that the fusion protein can be affinity  
CC purified on a nickel-agarose gel. Daunomycin C-14 hydroxylase  
CC is a P450-like enzyme capable of converting daunomycin to the  
CC anticancer agent doxorubicin. Host cells, especially Streptomyces  
CC host cells, transformed with pANT199 can be used in methods for  
CC the production of doxorubicin from daunomycin or for the  
CC hydroxylation and oxidation of other anthracyclines.  
SO Sequence 474 AA;

Query Match 7.38; Score 160; DB 1; Length 474;  
Best Local Similarity 23.68; Pred. No. 6.7e-08;  
Matches 96; Conservative 51; Mismatches 205; Indels 54; Gaps 14;

QY 39 QEAMAVIOESVNPDLVTRCNGHIAIRGQILREAYEDYRHFSE--CPFIPREAGEAY 96  
DB 83 REAGPVV-EVNAP-----AGSPAWVITDDALAREVLADPRFKGPDLPATAMRGVDDGL 135  
QY 97 D-----FIPTSMDPPEOROPALANQVGMVVDKLENRIQELACSLTESL-----R 143  
DB 136 DIPVELARPTFLIIVDGEDHRLKRIINAPNPRRLAERTDRIAIAIDRLITELADSSDR 195  
QY 144 POGQCNFTEDYAEPPPIRIFMLAGLP-----EEDIPHLKYL-----TDQMPRPD 188  
DB 196 SGEPAELIGCFAYHNPILLVICELGVPTDPMAREAVGLKALGLGPGSAGDGTDPA 255  
QY 189 GSMTFEAKKEALYDYLIIIEQOROKPGTDAISVANGVNGRPITSDAKRMGCLLVG 248  
DB 256 GDVPDTSALESU--LLEAVNAAARRKDRMTVRLYERAQAEFGSVSDQDLVYMTTGLIFA 313  
QY 249 GLDTPVNVNLSFSMEFLAKSPERQELTERPERIPACCELRLR-----FSLVADGRILTS 303  
DB 314 GHDTTGSFLGF--LLAEVLNAGRLAADADGALISFVEALRNHPVPYSL--WRFPAAT 367  
QY 304 DYEFHGVOIAKGGDQILLPQMLSLGLDERKNACPMHNVDFSRQKVSHTTGGSHLCLGQHIA 363  
DB 368 EVVIRGVRLPRGAPVLDIEGTNTDGRHNDARHAFNPRPSRRRLTFGDGPHVCIGEQOLA 427  
QY 364 RREIIVTLKEWLTIRPDSIA-PGAQIQ--HKSGLVSGVALPLVW 406  
DB 428 QLESRTMIGVLRSPQARLAIVPYEELRWCKRGAQTARLTDLR-VW 472

---

THIS PAGE BLANK (00118)

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 12:38:03 : Search time 66.92 seconds  
(without alignments)  
103.693 Million cell updates/sec

Title: US-09-246-451-11

Perfect score: 2180

Sequence: 1 TTTETISNNALAPLPPHVE.....IVSGVQALPLVMPDATTKAV 414

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 164575 seqs, 16761186 residues  
Total number of hits satisfying chosen parameters: 164575

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: Issued\_Patents\_AA.\*  
2: /cgn1\_7/ptodata/1/1aa/5A.COMB.pep.\*  
3: /cgn1\_7/ptodata/1/1aa/5B.COMB.pep.\*  
4: /cgn1\_7/ptodata/1/1aa/6.COMB.pep.\*  
5: /cgn1\_7/ptodata/1/1aa/6CCTUS.COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	365.5	16.8	416 3 US-09-320-878-18	Sequence 18, Appl
2	340	15.6	403 5 5212296-9	Patent No. 5212296
3	330	15.1	406 5 5212296-6	Patent No. 5212296
4	316	14.5	412 1 US-08-102-863-11	Sequence 11, Appl
5	316	14.5	412 4 PCT-US92-10885-11	Sequence 11, Appl
6	271.5	12.5	419 3 US-09-335-409-8	Sequence 8, Appl1
7	160	7.3	422 2 US-09-096-982-5	Sequence 5, Appl1
8	160	7.3	422 2 US-08-653-650A-5	Sequence 5, Appl1
9	160	7.3	422 2 US-09-096-982-8	Sequence 8, Appl1
10	160	7.3	422 2 US-08-653-650A-8	Sequence 8, Appl1
11	157	7.2	443 2 US-08-653-650A-9	Sequence 9, Appl1
12	157	7.2	443 2 US-08-396-218-2	Sequence 2, Appl1
13	155	7.1	422 1 US-08-760-116-2	Sequence 2, Appl1
14	155	7.1	422 1 US-08-194-981E-5	Sequence 7, Appl1
15	130	6.0	512 2 US-09-320-878-7	Sequence 7, Appl1
16	123.5	5.7	382 3 US-08-948-564-6	Sequence 6, Appl1
17	118	5.4	513 3 US-08-201-118-3	Sequence 3, Appl1
18	113.5	5.2	490 1 US-08-201-118-9	Sequence 9, Appl1
19	113.5	5.2	490 2 US-08-238-821B-9	Sequence 3, Appl1
20	113.5	5.2	490 2 US-08-238-821B-9	Sequence 3, Appl1
21	113.5	5.2	490 4 PCT-US95-05744-3	Sequence 3, Appl1
22	113.5	5.2	490 4 PCT-US95-05744-9	Sequence 9, Appl1
23	113.5	5.2	501 3 US-08-906-791-2	Sequence 2, Appl1
24	113	5.1	490 1 US-08-201-118-7	Sequence 7, Appl1
25	111.5	5.1	490 2 PCT-US95-05744-7	Sequence 7, Appl1
26	111.5	5.1	504 1 US-08-457-274A-25	Sequence 25, Appl1
27	111.5	5.1		
28	111.5	5.1		

29	111.5	5.1	504 4 PCT-US95-05758-25	Sequence 25, Appl
30	110	5.0	492 3 US-08-724-466B-2	Sequence 2, Appl
31	105.5	4.8	496 1 US-08-313-075A-50	Sequence 50, Appl
32	104	4.8	497 3 US-08-724-466B-4	Sequence 4, Appl1
33	102.5	4.7	472 2 US-08-622-166A-2	Sequence 2, Appl1
34	102.5	4.7	472 2 US-08-622-166A-4	Sequence 4, Appl1
35	99.5	4.6	490 1 US-08-201-118-1	Sequence 1, Appl1
36	99.5	4.6	490 2 US-08-238-821B-1	Sequence 1, Appl1
37	99.5	4.6	490 4 PCT-US95-05744-1	Sequence 1, Appl1
38	99.5	4.6	498 1 US-08-457-274A-24	Sequence 24, Appl
39	99.5	4.6	498 4 PCT-US95-05758-24	Sequence 24, Appl
40	96.5	4.4	500 2 US-08-314-601-2	Sequence 2, Appl1
41	96.5	4.4	500 4 PCT-US95-13051-2	Sequence 2, Appl1
42	94.5	4.3	523 3 US-08-606-505B-67	Sequence 67, Appl
43	94.5	4.3	1026 1 US-07-998-003A-95	Sequence 95, Appl
44	94.5	4.3	1026 1 US-08-453-274B-95	Sequence 95, Appl
45	94.5	4.3	1026 1 US-08-453-695A-95	Sequence 95, Appl

ALIGNMENTS

```
RESULT 1
US-09-320-878-18
: Sequence 18, Application US/09320878A
: Patent No. 6117659
:
: GENERAL INFORMATION:
: APPLICANT: ASHLEY, Gary
: APPLICANT: BETLACH, Melanie C.
: APPLICANT: BETLACH, Mary C.
: APPLICANT: MCDANIEL, Robert
: APPLICANT: TANG, Li
:
: TITLE OF INVENTION: RECOMBINANT NARONOLIDE POLYKETIDE SYNTHASE
: FILE REFERENCE: 300622002120
: CURRENT APPLICATION NUMBER: US/09/320, 878A
: EARLIER FILING DATE: 1999-05-27
: EARLIER APPLICATION NUMBER: CIP OF 09/141,908
: EARLIER FILING DATE: 1998-08-28
: EARLIER APPLICATION NUMBER: CIP OF 09/073,538
: EARLIER FILING DATE: 1998-05-06
: EARLIER APPLICATION NUMBER: CIP OF 08/846,247
: EARLIER FILING DATE: 1997-04-30
: EARLIER APPLICATION NUMBER: 60/119,139
: EARLIER FILING DATE: 1999-02-08
: EARLIER APPLICATION NUMBER: 60/100,880
: EARLIER FILING DATE: 1998-09-22
: EARLIER APPLICATION NUMBER: 60/087,080
: EARLIER FILING DATE: 1998-05-28
: NUMBER OF SEQ ID NOS: 34
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 18
: LENGTH: 416
: TYPE: PRT
: ORGANISM: Streptomyces venezuelae
US-09-320-878-18
```

Query Match 16.8%, Score 365.5; DB 3; Length 416;  
Best Local Similarity 28.3%; Pred. No. 8e-30;  
Matches 97; Conservative 62; Mismatches 163; Indels 21; Gaps 7;

QY	76	EDYRHFSSCEPIPEAGCAVDIFTSMDPEBQRFALANOVGMVYDKLENRIQELA	135
DB	72	KDWR--NSTPLTEAEALNHNMLESS--DPRHRTLRKLVAREFTMRVVELLRPRVQIV	127
QY	136	CSLIESSL--RPGGOCNFEDEVAEPPRIRIFMLGLTPREDIPHIKYLTDOQTRPDSMTF	193
DB	128	DGLVDAKMLAAPDGRADLMESLAWPLITVISELGVPEPDRAAFVWTDAFVPDDPAQA	187
QY	194	AEAKKALVDYLIPIIEORRPGTDAIS--IVANGVNGRPITSDEAKRMCCILLVGIDT	252
DB	188	QTANAMGSGYLSRLIDSKRGDGDGLLSALVRTSDEDSRLTSEELGMHILLVAGHET	247

OY 253 VNFPSFMEFLAKPESRQELIESPERIPACCELLFRFSIVAGRIITSDYE----- 307  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db 248 TVNLIANGVALLSHPPDOLALRADMTLIDGAVEMLR-----YGPVESATRYFPVEFY 302  
  
OY 308 --HGVOALKKDOILLPOMLSGLDERKNACPMHVDFSRKVSHTEFGHSHLCIGOHLAR 365  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db 303 DLDGVIAICAGPTVLAVLVADAHRTPRFPDPHPFDLRPDTAGHLACHGHHCFCIAPLARL 362  
  
OY 366 ELVTLEKEMTRIPDFS--IAPGAIOHKSGISVGVALPLTW 406  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db 363 EARIVARALLERCPLDALDVSPGELWTPNPWTIRGLKALPIRW 405

RESULT 2  
5212296-9  
Patent No. 5212296  
APPLICANT: DEAN, CAROLINE; HARDER, PATRICIA A.; LETO, KENNETH  
J.; O'KEEFE, DANIEL P.; OMER, CHARLES A.; ROMESSER, JAMES A.  
; TEPPERMAN, JAMES M.

TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING CYTOCHROMES

NUMBER OF SEQUENCES: 19  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/569,781  
FILING DATE: 23-AUG-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 464,499  
FILING DATE: 12-JAN-1990  
APPLICATION NUMBER: 405,605  
FILING DATE: 11-SEP-1989  
SEQ ID NO.:9  
LENGTH: 403  
5212296-9

	Query Match	15.6%;	Score 340; DB 5; Length 403; Best Local Similarity 28.9%; Pred. No. 3.5e-27; Matches 105; Conservative 63; Mismatches 165; Indels 30; Gaps 13.
OY	63 WATRGQILREAYEDRYHFSSEC-----PFIPEAGAEVDFIPT--SMDPRBQRFRALA	115	:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Dd	52 WLVTRHODVRRAVLGDPR-FSADAHRTGPFPLTAGCREIIGTPFLRMDDPENARLRML	110	:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
OY	116 NQVVGMPPVDLENNIOELACSLIESLAP-GQGCFPTDYEAPPFIRIFMLAGIPEDI	174	:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Dd	111 TADELVKKVEAMREPVOALADDKRRMTTGKSASAVTEFALPDLSVICLLGVYEDH	170	:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
OY	175 PHL----KYTLDMOTRPDGSMTFEAKBALDYDLPIITEORRKGPGTAIS-IIVANGOVN	229	:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Dd	171 AFQERSRHVLTLTRSTPE---EVRAAOELLELEYLARLKARTKRERDDAIISRLVARGEID	227	:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
OY	230 GRPTSDKAEMCGILLVGGLDTYNVFNSFEKLAKSPENHOELIERPERIPAACEELL	289	:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Dd	228 DTQTAT----MGRIILAAGHETTANMALSTLYLIINPDOLARIARAEPALKVGAEEII	282	:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
OY	290 RRESLVA DG-RILTSDVEFHVOLVKKDQLLPOMLSGLDERKNACP--MHVDSRKCV	345	:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Dd	283 RYLITVNHGVSRIATEDVILIGRTIAAEGVC--MISSANDALEFPGGDDLVIARDAAR	340	:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
OY	346 SHHTFGSHCLGOHLARETIITLKEWLTRIPDEISA-PGAOIQHSGT-VSCVOALP	403	:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Dd	341 RHVAFGFVHCLOCPLEARVELQIMAITELLRLPOLRLAVALPHETIPFGDMAITYGHSP	400	:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
OY	404 LVW 406		:
Dd	401 IAW 403		:
<b>RESULT      3</b>			
Patent NO. 5212296-6			
APPLICANT: DEAN CAROLINE HARDER PATRICIA A.; LETO KENNETH J.; O'KEEFE DANIEL P.; OMER CHARLES A.; ROMSSSER JAMES A.			

```

TEPPERMAN, JAMES M.
  TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING
CYTOCHROMES
  NUMBER OF SEQUENCES: 19
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/07/569,781
    FILING DATE: 23-AUG-1990
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: 464,499
    FILING DATE: 12-JAN-1990
    APPLICATION NUMBER: 405,605
    FILING DATE: 11-SEP-1989
  SEQ ID NO.: 6
  LENGTH: 406
5212296-6

```

QY	31	PSNLSHGQV--EAAVAFLDSSNVPLWVRCNGH-WLAIKQLLREAEIQR-----	179
Db	17	PSNNSCPYLPDGYAOLRDFPGPLHRTVLYDGRQAAVVTFKHAAKLLGPRLLSNRTDO	76
QY	80	HESSECPPII--RAGEAYEDTPIISMPPRQROCFRALANOVGMPPVYDKLENRQELACS	137
Db	77	NFPATSPFEAFVRSPOAF---IGLDPBEGHTRRMTTISEFTYKRIKGRAPVEEIVHG	132
QY	138	LIESLRPOG--QCNTEDYAEFPFIRIMLLAGLPEDDIPHLKYLTDDMTPRDGSMPFAA	166
Db	133	FLIDMLANGPPADLVSOFAIDPVSMTVICRLGVYADHEFFQDASKRLVSTPDQSAITA	132
QY	197	KEALYDYLPIIEOROKPGTDAI--SIYANGVNGRPITSDCAKRMGKLLVGLDTPVYN	255
Db	193	RNDLAGTYLDIGLTQFQIERPGAGLVGALVAQDLANGE--IDREELISTAMLLLIAGHETTAS	251
QY	256	FLFSMEFLAKSPENROELIERPERIPACEELLRRSL--VADGRILTSDYEPHFVOLK	313
Db	252	MTSLSVTLLDHPQYAAALRADRSFLYGVAYEELLRYLATINDIAGGRVATDIEVGHILIR	311
QY	314	KGDQILLPQMLISGIDERRKMACPMHVDPSRQKVSHTTGGSHLCLGQHLARREIYVLTKE	373
Db	312	AGEGVIVANNSIANDGTVYVEDPDALDIHRSARHHLAAGFVHCQGLQNLARLETEVILNA	371
QY	374	WLTRIPDESTA--PGAQIQHKSG--IVSQVALPLTW	406
Db	372	LMDRVPTLRILAVPEQLVLRGTTIGQVNEPLVTW	406

RESULT 4  
 US-08-102-863-11  
 : Sequence 11, Application US/08102863  
 : Patent No 5465580  
 :  
 : GENERAL INFORMATION:  
 : APPLICANT: SARIASLANI, SIMA  
 : TITLE OF INVENTION: CONSTITUTIVE  
 : TITLE OF INVENTION: EXPRESSION OF P450C50Y  
 : TITLE OF INVENTION: AND FERREDOXIN-SOY IN  
 : TITLE OF INVENTION: STREPTOMYCES  
 : NUMBER OF SEQUENCES: 12  
 :  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSSEE: E. I. DU PONT DE NEMOURS  
 : ADDRESSSEE: AND COMPANY  
 : STREET: 1007 MARKET STREET  
 : CITY: WILMINGTON  
 : STATE: DELAWARE  
 : COUNTRY: USA  
 : ZIP: 19898  
 :  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS

```

SOFTWARE: Patent Release #1.0',
SOFTWARE: Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/102,863
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/807,000
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: GALLEGOS, R. THOMAS
REGISTRATION NUMBER: 32,692
REFERENCE/DOCKET NUMBER: CR-9000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-892-7342
TELEFAX: 302-892-7949
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 412 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-102-863-11

```

Query Match	14.5%;	Score 316;	DB 1;	Length 412;
Best Local Similarity	27.2%;	Pred. NO. 1.2e-24;		
Matches	84;	Conservative	56;	Mismatches 163;
			Indels	6;
			Gaps	5;

QY	103	MDPEQOQFRLAQVUGMVPVDLEJRIERIOELACSLIESLJRPQG-OCNFEDVAEPPR	161
Db	105	VDDPEHNTQRMRLPFPSVKRIGALRRRIGIETUYDRILLDAMERGSRPRLVSAFALP	164
QY	162	IFMILLAGPREDDIHNKLYDDQMTFRPGSMTFAEKALYDYLPIITEQROKRGTDIAIS	221
Db	165	VICALGVPADNAHFEERSORILLRGSGADVUNBARLEETLYCALIDRRKRAEGRDOLD	224
QY	222	IVANGQVNGRPTSDSEKRMGGLLVGSLDTUVNLFJFSMEFLAKSPENHOELTERPERI	281
Db	225	ELIHRDHPDGVOREQLVARAVILLINGHETNAMISLGFITLLSHEDQALALRAGST	284
QY	282	PACEELLRRPSLVADG--RLISDYFNGVQLKGGQILLPOMLSGLDERKNAQCPMHVD	339
Db	285	AVUYEELL-RLSLAEGLORLATEDEMDVDQATIKGGVVFSTSLIRMDADVPRARETLD	343
QY	340	FSRQKVSHTTEGHSNLSLGLNARREIITYTLKEMILRIDEFSIA-PGAOIQHKSQ-IVS	397
Db	344	MDPRARHHLAFGFEVHCISQONLARALDDIAMRTLPFRLLGRLAVRANHRIKRPQTIQ	403
QY	398	GVQALPLVW	406
Db	404	GLLDPLVAM	412

RESULT 5  
PCT-US92-10885-11  
Sequence 11, Application PC/TUS9210885  
GENERAL INFORMATION:  
APPLICANT: SARIASLANI, SIMA  
TITLE OF INVENTION: CONSTITUTIVE  
TITLE OF INVENTION: EXPRESSION OF P450SOY  
TITLE OF INVENTION: AND FERREDOXIN-SOY IN  
STREPTOMYCES  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: E. I. DU PONT DE NEMOURS  
ADDRESS: E. I. DU PONT DE NEMOURS  
STREET: 1007 MARKET STREET  
CITY: WILMINGTON  
STATE: DELAWARE  
COUNTRY: USA  
ZIP: 19898

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch,
MEDIUM TYPE: 1.0 MB
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh System, 6.0
SOFTWARE: Microsoft Word, 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/10885
FILING DATE: 19921216
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: GALEGOS, R. THOMAS
REGISTRATION NUMBER: 32,692
REFERENCE/DOCKET NUMBER: CR-9000-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-892-7342
TELEFAX: 302-892-7949
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 412 amino acids
TYPE: AMINO ACID
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
PCT-US92-10885-11

```

Query Match	14.5%;	Score 316;	DB 4;	Length 412;
Best Local Similarity	27.2%;	Pred. No. 1.2e-24;		
Matches 84;	Conservative 56;	Mismatches 163;	Indels 6;	Gaps 5;

[illegible]

RESULT 6  
 US-09-335-409-8  
 : Sequence 8, Application US/09335409  
 : Patent No. 6121029  
 : GENERAL INFORMATION:  
 : APPLICANT: Schupp, Thomas  
 : APPLICANT: Ligon, James  
 : APPLICANT: Molnar, Istvan  
 : APPLICANT: Zirkle, Ross  
 : APPLICANT: Cyr, Devon  
 : APPLICANT: Goelach, Joern  
 : TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
 : FILE REFERENCE: 4-30582A  
 : CURRENT APPLICATION NUMBER: US/09/335,409  
 : CURRENT FILING DATE: 1999-06-17  
 : NUMBER OF SEQ ID NOS: 30  
 : SOFTWARE: PatentIn ver. 2.0





```

: NAME: Gollick, Mary E.
: REGISTRATION NUMBER: 34829
: REFERENCE/DOCKET NUMBER: 22727/00131
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 216-622-8458
: TELEFAX: 216-241-0816
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 422 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-653-650A-5
```

Query Match 7.3%; Score 160; DB 2; Length 422;

Best Local Similarity 23.6%; Pred. No. 2.3e-08;

Matches 96; Conservative 51; Mismatches 205; Indels 54; Gaps 14;

```

QY 39 QEAMAVLQESNVPLVWTRCNGHWIATRGQILREAYEDYRHFSSE--CPFLPREGAY 96
D 31 REAGVYV-EVAP-----AGSPAVITDDALAREVLADPFVKGPDLPAMRGVDGL 83
QY 97 D-----FIPTSMDPEEQRFALANQVGMVYDKLENRIQELACSLIEST-----R 143
D 84 DIPVELRPFTLIAVDGSDHRLRIHAPFNPRLAERTDIRIAIAIDRLLELADSSDR 143
QY 144 POGOCNFTEDEAEPPIRIFMLAGLP-----EEDIPHLKYL-----TDQMTRPD 188
D 144 SGEPAELIGGFAHYHPLLVICELGVPTDPMAREAVGLKALIGSPQSGAGDTPA 203
QY 189 GSMFPAEAKKALYDLPIIEORROKPGTDAISIVANGVNGRPITSDPAKRMCGLLVG 248
D 204 GDVPTSALESF--LLEAVHARRKDTMTMTVLTERAQAEGSVSDQLVYMTGLIFA 261
QY 249 GLDTVYNELFSMEFLAKSPERHOLLERPERIPACCELLR-----FSLVADRILTS 303
D 262 GHDTGSFLGF--LLAEVLGRLAADADGDAISRFVEALRNHPVPYST--WRPAAT 315
QY 304 DYEFHGVOLKKGDQILLPOMLSGDERKNACPMHVDFSROKVSHTTFGSHLCLGONHA 363
D 316 EVYIRGVRLPRGARVYLDIEGTNTDGRHNDAPNAPHNRPSPRRLLTFDGRHYCIGBOLA 375
QY 364 RREIIVLKEMLTIRPDFSIA-PCAQIQ--HKSGIVSGVALPLVW 406
D 376 QLESRTMIGVLRSPQARLAVPYEELRMCKRGAQTARTLDLP-VW 420
```

RESULT 9

US-09-096-982-8

; Sequence 8, Application US/09096982

; Patent No. 5962293

; GENERAL INFORMATION:

; APPLICANT: Strohl, William R.

; APPLICANT: Dickens, Michael L.

; APPLICANT: Desanti, Charles L.

; TITLE OF INVENTION: METHOD OF PRODUCING DOXORUBICIN

; NUMBER OF SEQUENCES: 9

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CALFEE, HALTER & GRISWOLD

; STREET: 800 Superior Avenue, Suite 1400

; CITY: Cleveland

; STATE: Ohio

; COUNTRY: USA

; ZIP: 44114-2688

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/096,982

; FILING DATE:

```

: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Gollick, Mary E.
: REGISTRATION NUMBER: 34829
: REFERENCE/DOCKET NUMBER: 22727/00131
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 216-622-8458
: TELEFAX: 216-241-0816
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 474 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-09-096-982-8.
```

Query Match 7.3%; Score 160; DB 2; Length 474;

Best Local Similarity 23.6%; Pred. No. 2.8e-08;

Matches 96; Conservative 51; Mismatches 205; Indels 54; Gaps 14;

```

QY 39 QEAMAVLQESNVPLVWTRCNGHWIATRGQILREAYEDYRHFSSE--CPFLPREGAY 96
D 83 REAGVYV-EVAP-----AGSPAVITDDALAREVLADPFVKGPDLPAMRGVDGL 135
QY 97 D-----FIPTSMDPEEQRFALANQVGMVYDKLENRIQELACSLIEST-----R 143
D 136 DIPVELRPFTLIAVDGSDHRLRIHAPFNPRLAERTDIRIAIAIDRLLELADSSDR 195
QY 144 POGOCNFTEDEAEPPIRIFMLAGLP-----EEDIPHLKYL-----TDQMTRPD 188
D 196 SGEPAELIGGFAHYHPLLVICELGVPTDPMAREAVGLKALIGSPQSGAGDTPA 255
QY 189 GSMFPAEAKKALYDLPIIEORROKPGTDAISIVANGVNGRPITSDPAKRMCGLLVG 248
D 256 GDVPTSALESF--LLEAVHARRKDTMTMTVLTERAQAEGSVSDQLVYMTGLIFA 313
QY 249 GLDTVYNELFSMEFLAKSPERHOLLERPERIPACCELLR-----FSLVADRILTS 303
D 314 GHDTGSFLGF--LLAEVLGRLAADADGDAISRFVEALRNHPVPYST--WRPAAT 367
QY 304 DYEFHGVOLKKGDQILLPOMLSGDERKNACPMHVDFSROKVSHTTFGSHLCLGONHA 363
D 368 EVYIRGVRLPRGARVYLDIEGTNTDGRHNDAPNAPHNRPSPRRLLTFDGRHYCIGBOLA 427
QY 364 RREIIVLKEMLTIRPDFSIA-PCAQIQ--HKSGIVSGVALPLVW 406
D 428 QLESRTMIGVLRSPQARLAVPYEELRMCKRGAQTARTLDLP-VW 472
```

RESULT 10

US-08-653-650A-8

; Sequence 8, Application US/08653650A

; Patent No. 5976830

; GENERAL INFORMATION:

; APPLICANT: Strohl, William R.

; APPLICANT: Dickens, Michael L.

; TITLE OF INVENTION: METHOD OF PRODUCING DOXORUBICIN

; NUMBER OF SEQUENCES: 9

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CALFEE, HALTER & GRISWOLD

; STREET: 800 Superior Avenue, Suite 1400

; CITY: Cleveland

; STATE: Ohio

; COUNTRY: USA

; ZIP: 44114-2688

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:



MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/653,650A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Goldick, Mary E.  
REGISTRATION NUMBER: 34829  
REFERENCE/DOCKET NUMBER: 22727/00131  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 216-241-0816  
TELEFAX: 216-241-0816  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 443 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-653-650A-9

Query Match 7.2%, Score 157; DB 2; Length 443;  
Best Local Similarity 23.6%, Pred. No. 5,2e-08;  
Matches 95; Conservative 50; Mismatches 210; Indels 48; Gaps 13;

QY 39 QEAMAVLQESNVPDLVWTRCNGHWIATRGOLIREAYEDYRHFSSE--CPETPREAGEAY 96  
DB 52 REAGFVV--EVNAP-----ACGPAAVITDDALAREVLADPRVKPDLAPRAMRGVDGL 104  
QY 97 D-----FIPTSMDPREQOFRALANQVGMRYVDKLENIOELASLIESL-----R 143  
DB 105 DIPVELRPFTLLIADGSDHRRRLRIHAPAFNPRLAERTORIAIAIADRLTELDSSDR 164  
QY 144 POGOCNFTEDYAEPRIRIFMLAGLP-----EEDIPHLKYL-----TDQMTPRD 188  
DB 165 SGERALLEGFAVHFRPLVIGELGVPTDRPMAREAVGLKALDGGPQSGSGTTPRA 224  
QY 189 GSMFAEAKKALYDYLPIIEORROKPGTDAISIVANGQVNGRPITSDANKMCGLLVG 248  
DB 225 GDVPTSALESF--LLEAVHAAARKDRTMTRVLYERAQAEFGVSDDQLVYMTGLIFA 282  
QY 249 GUDTVVNLFSMERFLASPREHRELIERPERIPACCELLRRSLV--ADGRILTSYE 306  
DB 283 GDDTGSFLGF--LLAEVLAGRLAADADGALISRFVEEALKHNPVYTYLMRFATEV 339  
QY 307 FHGVOLKKDQDILLPQMLSGIDERKNACPMHVDFSROKVSHTFEGSHLCLGHLARE 366  
DB 340 IRGVALPRGAVLVLDIEGTINDGRHNDARHAFHPRPSRRRLTFEGDGHYICGDLADE 399  
QY 367 IIVTLKEMLTIRPDFSIA--PGAQIO--HKSGIVSGVQALPLVW 406  
DB 400 SRTMIGVLRSPQARLAVPYEELRWCKGAQTARLTDLR--VW 441

RESULT 13  
US-08-396-218-2  
; Sequence 2, Application US/08396218  
; Patent No. 5695966  
; GENERAL INFORMATION:  
; APPLICANT: INVENTI, Augusto  
; APPLICANT: BREME, Umberto  
; APPLICANT: COLOMBO, Anna L  
; APPLICANT: HUTCHINSON, Charles R  
; APPLICANT: OTTEN, Sharee  
; APPLICANT: SCOTTI, Claudio  
; TITLE OF INVENTION: PROCESS FOR PREPARING DOXORUBICIN  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIKAIIDO, MARWELSTEIN, MURRAY & ORAM

STREET: 655 Fifteenth Street, N. W., Suite 330 - G  
; STREET: Street Lobby  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005-5701  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/396,218  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: KITTS, Monica C  
REGISTRATION NUMBER: 36,105  
REFERENCE/DOCKET NUMBER: P1615-5002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202/638-5000  
TELEFAX: 202/638-4810  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 422 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-396-218-2

Query Match 7.1%, Score 155; DB 1; Length 422;  
Best Local Similarity 23.8%, Pred. No. 7.8e-08;  
Matches 97; Conservative 51; Mismatches 202; Indels 58; Gaps 15;

QY 39 QEAMAVLQESNVPDLVWTRCNGHWIATRGOLIREAYEDYRHFSSECPETPREAGEAYDF 98  
DB 31 REAGFVV--EVNAP-----ACGPAAVITDDALAREVLADPRVKD-----PDLAPAAARG 78  
QY 99 IPTSMD--PREQORFALA-----NQVGMRYVD-----KLENIOELASLIESL- 142  
DB 79 VDDGIDIPVELRPFTLLIADGSEAHRRRLRIHAPAFNPRLAERTORIAIAGRLTEIA 138  
QY 143 ---RPOGOCNFTEDYAEPRIRIFMLAGLP-----EEDIPHLKYL-----TDQ 183  
DB 139 DASGRSGKPAELIGFAHFRPLVIGELGVPTDRPMAREAVGLKALDGGPQSGGSD 198  
QY 184 MTRPDGSMFAEAKKALYDYLPIIEORROKPGTDAISIVANGQVNGRPITSDANKMCG 243  
DB 199 GTDPAGVPRDTSALBSL--LLEAVHSARRNDPTMTRVLYERAQAEFGVSDDQLVYMIT 256  
QY 244 LILVGLDITVYNFLFSMERFLASPREHRELIERPERIPACCELLRRPSLV--ADGRIL 301  
DB 257 GLIFAGHDTTGSFLGF--LLAEVLAGRLAADADDAVSREFEALRHNPRVYTYLMRFA 313  
QY 302 TSDYFHGVOLKKDQDILLPQMLSGIDERKNACPMHVDFSROKVSHTFEGSHLCLGHL 361  
DB 314 ATEVTIGVLRPRGAVLVLDIEGTINDGRHNDARHAFHPRPSRRRLTFEGDGHYICG 373  
QY 362 LARREIIVTLKEMLTIRPDFSIA--PGAQIO--HKSGIVSGVQALPLVW 406  
DB 374 LAQLESRTMIGVLRSPQARLAVPYEELRWCKGAQTARLTDLR--VW 420

RESULT 14  
US-08-760-116-2  
; Sequence 2, Application US/08760116  
; Patent No. 5786190  
; GENERAL INFORMATION:  
; APPLICANT: INVENTI, Augusto  
; APPLICANT: BREME, Umberto  
; APPLICANT: COLOMBO, Anna L  
; APPLICANT: HUTCHINSON, Charles R



Db 365 PYMAFILETFRHSSEVPETIPHSSTRDTSIKGEYIPKRCVFVNOMQINHDOKLWNP 424  
OY 336 -----MHVDFSRKV---SHTFGHSHLCLGOHLARREIIVTLKEMLTRIPDPSIAP 385  
Db 425 EFLPERFLPDGAIKVLSEKVIIFGMRKCIGETVARWEVFLFLAILLORV-EFSVPL 483  
OY 386 GAOI 389  
Db 484 GVKV 487

Search completed: October 4, 2000, 12:38:05  
Job time: 21422 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 12:39:34 : Search time 79.83 Seconds  
(without alignments)  
359.568 Million cell updates/sec

Title: US-09-246-451-11  
Perfect score: 2180  
Sequence: 1 TTEIQTNNALPLPPHVE.....IVSGVALPLVMDPATTKAV 414

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: SPREMBL\_12:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_podent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	365.5	16.8	416	2	087605 streptomyc
2	364	16.7	396	2	052544 amycolatops
3	364	16.7	397	2	059523 micromonos
4	347	15.9	410	2	09x5p9 streptomyc
5	346.5	15.9	404	2	054302 streptomyc
6	344.5	15.8	407	2	059819 streptomyc
7	329	15.1	411	2	032460 actinomadu
8	326	15.0	406	2	059723 pseudomon
9	325.5	14.9	420	2	092H01 streptomyc
10	324.5	14.9	375	2	031785 bacillus su
11	321	14.7	407	2	09x5p8 streptomyc
12	318	14.6	310	2	032927 mycobacteri
13	317	14.5	406	2	087192 streptomyc
14	316	14.5	410	2	09x5p7 streptomyc
15	313.5	14.4	410	2	059831 streptomyc
16	312	14.3	400	2	09x5p0 mycobacteri
17	303.5	13.9	397	2	024727 nocardioide
18	303	13.9	417	2	059910 streptomyc
19	301	13.8	406	2	085697 streptomyc

20	294.5	13.5	337	2	052816 amycolatops
21	293	13.4	351	2	052572 amycolatops
22	291.5	13.4	406	2	087675 amycolatops
23	291	13.3	388	2	p77977 streptomyc
24	279	12.8	421	2	052561 amycolatops
25	277.5	12.7	310	2	p96562 amycolatops
26	267	12.2	398	2	087674 amycolatops
27	265.5	12.2	437	2	092FC0 mycobacteri
28	264	12.1	403	2	052560 amycolatops
29	262.5	12.0	411	2	09x803 streptomyc
30	261.5	12.0	386	2	059921 streptomyc
31	261	12.0	395	2	09x5f2 streptomyc
32	256	11.7	426	2	092FC3 mycobacteri
33	255	11.7	406	2	052823 amycolatops
34	250	11.5	411	2	060005 putative
35	248.5	11.4	511	2	052569 amycolatops
36	239	11.0	433	2	09x5C6 streptomyc
37	226	10.4	391	2	052822 amycolatops
38	225	10.3	398	2	059079 amycolata a
39	224.5	10.3	391	2	087673 amycolatops
40	215	9.9	397	2	052802 amycolatops
41	203.5	9.3	419	2	050242 agrobacteri
42	182.5	8.4	119	2	085655 streptomyc
43	179	8.2	120	2	085653 streptococ
44	174.5	8.0	313	2	09x418 myxococcus
45	164	7.5	120	2	085650 streptomyc

## ALIGNMENTS

RESULT 1  
ID 087605 PRELIMINARY: PRT: 416 AA.  
AC 087605;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
DE CYTOCHROME P450 MONOOXYGENASE.  
GN PICK OR PICK.  
OS Streptomyces violaceus (Streptomyces venezuelae), and  
OS Streptomyces venezuelae.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC15439;  
RA BETLACH M.C., KEALEY J.T., BETLACH M.C., ASHLEY G.W., MCDANIEL R.;  
RT "Characterization of the macrolide P450 hydroxylase from Streptomyces  
RT venezuelae which converts narboxycin to ploxycin.",  
RL Biochemistry 0:0-0(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES-S.VENEZUELA; STRAIN-ATCC15439;  
RA MEDLINE: 99051447.  
XUE Y., WILSON D., ZHAO L., LIU H.-W., SHERMAN D.H.;  
RT "Hydroxylation of macrolactones VC-17 and narboxycin is mediated by  
RT the P450 encoded cytochrome P450 in Streptomyces venezuelae.",  
RL Chem. Biol. 5:661-669(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES-S.VENEZUELA; STRAIN-ATCC15439;  
RA MEDLINE: 98445333.  
XUE Y., ZHAO L., LIU H.-W., SHERMAN D.H.;  
RT "A gene cluster for macrolide antibiotic biosynthesis in Streptomyces  
RT venezuelae: architecture of metabolic diversity.",  
RL Proc. Natl. Acad. Sci. U.S.A. 95:12111-12116(1998).  
DR EMBL: AF087022; AAC64105.1; -;  
DR EMBL: AF079139; AAC68886.1; -;  
DR HSSP: Q00441; LOXA.  
DR PFAM: PF00067; P450; 1.  
DR PRINTS: PR00359; BP450.  
KW Monooxygenase.





QY	224	ANGOVNAGRITSD	EAKRRCGILLVGGIDTVN	FNFSMEFLAKSPENHOELIERE	PRIPA	283																																													
DB	210	VOARQOOSLS <td>EQELLDAIGLAVG<th>STTTQIAD</th><td>EVLMTRELRQLDRPELLPS</td><td>269</td></td>	EQELLDAIGLAVG <th>STTTQIAD</th> <td>EVLMTRELRQLDRPELLPS</td> <td>269</td>	STTTQIAD	EVLMTRELRQLDRPELLPS	269																																													
QY	284	ACEELLRRFSL	--VADGRILTSDE <td>FEHGVOLKKGDQILLPOMLSGLDERKNACPMH</td> <td>NDF</td> <td>340</td>	FEHGVOLKKGDQILLPOMLSGLDERKNACPMH	NDF	340																																													
DB	270	AVEELLRWPL <td>GVGTAFPRVAV<td>EDVTINGVTLIRAE<td>PEVLASGANNRQAO</td><td>PRDADRIV 329</td></td></td>	GVGTAFPRVAV <td>EDVTINGVTLIRAE<td>PEVLASGANNRQAO</td><td>PRDADRIV 329</td></td>	EDVTINGVTLIRAE <td>PEVLASGANNRQAO</td> <td>PRDADRIV 329</td>	PEVLASGANNRQAO	PRDADRIV 329																																													
QY	341	SNOKSHTT <td>FGHSLCIGOH<td>LAREITVTLK<td>WTLTRIPDSIA</td><td>-PQAOIQHKS</td><td>-IVSG 398</td></td></td>	FGHSLCIGOH <td>LAREITVTLK<td>WTLTRIPDSIA</td><td>-PQAOIQHKS</td><td>-IVSG 398</td></td>	LAREITVTLK <td>WTLTRIPDSIA</td> <td>-PQAOIQHKS</td> <td>-IVSG 398</td>	WTLTRIPDSIA	-PQAOIQHKS	-IVSG 398																																												
DB	330	DTRPNQHG <td>FGVGHHC<td>GAPLAREV<td>LOVALEVLRLR<td>PGIRLGIPTOLR<td>MSGMLRG 389</td></td></td></td></td>	FGVGHHC <td>GAPLAREV<td>LOVALEVLRLR<td>PGIRLGIPTOLR<td>MSGMLRG 389</td></td></td></td>	GAPLAREV <td>LOVALEVLRLR<td>PGIRLGIPTOLR<td>MSGMLRG 389</td></td></td>	LOVALEVLRLR <td>PGIRLGIPTOLR<td>MSGMLRG 389</td></td>	PGIRLGIPTOLR <td>MSGMLRG 389</td>	MSGMLRG 389																																												
QY	399	VOALPLW	406																																																
DB	390	PLELPVW	397																																																
RESULT	4																																																		
Q9X5P9	ID	Q9X5P9	PRELIMINARY;	PRT:	410 AA.																																														
AC	Q9X5P9																																																		
DT	01-NOV-1999	(TREMblrel. 12, Created)																																																	
DT	01-NOV-1999	(TREMblrel. 12, Last sequence update)																																																	
DT	01-NOV-1999	(TREMblrel. 12, last annotation update)																																																	
DE	CYCHROME	P450 HYDROXYLASE ORF3.																																																	
OS	Streptomyces	lavendulae																																																	
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;																																																		
OC	Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.																																																		
OC	(1)																																																		
RN	SEQUENCE FROM N.A.																																																		
RP	STRAIN-NRRL 2564;																																																		
RC	MEDLINE: 99201491.																																																		
RX	MAO Y.Q., VAROGU M., SHERMAN D.H.;																																																		
RA	"Molecular characterization and analysis of the biosynthetic gene																																																		
RT	cluster for the antitumor antibiotic mitomycin C from Streptomyces																																																		
RT	lavendulae NRRL 2564."																																																		
RL	Chem. Biol. 6:251-263(1999).																																																		
DR	EMBL: AF127374; AAD28449.1;																																																		
SO	SEQUENCE 410 AA; 45887 MW; 507C7F38 CRC32;																																																		
Query Match	15.9%;	Score 347;	DB 2;	Length 410;																																															
Best Local Similarity	27.0%;	Pred. No. 2.2e-20;																																																	
Matches 113;	Conservative 69;	Mismatches 193;	Indels 44;	Gaps 14;																																															
QY	9	ANLADLP	-HVEPHLV	PDFDMY	NPSNL	SGVOEAMV	LDQSNV	PDLV	WTRCNG	GHMIATR 67																																									
DB	15	AGEA	APFPHADRL	-EPDPY	-----	WEPLR	REPR	LO	RVTL	PLYGSE	AMLATR 60																																								
QY	68	GOL	REARE	EDVRH	SESE	---CPPI	REAGE	AVDF	IP	-----	TSMOP	PE	RO	RALAN 116																																					
DB	61	YOD	NA	VA	FEAD	-RRSR	OLA	VA	GAPR	-----	FLPH	OP	PP	DA	VL	SV	GPCD	AR	LR 112																																
QY	117	QV	NG	PV	VD	KLE	NR	LI	ES	LA	CS	IS	LE	SR	EOG	-OCN	TE	DE	YA	EE	FP	IR	IM	LA	GL	PE	ED	IP 175																							
DB	113	KVF	PR	RV	ED	KR	PL	Q	FR	AD	GL	DA	ME	EM	GR	PA	DL	VE	D	SP	UL	FA	NS	MC	EL	LG	V	PE	DRK 172																						
QY	176	HL	KY	IT	DO	M	TR	PD	GS	MT	FA	E	A	K	E	---AL	YD	LI	PI	EO	R	OK	PG	TD	AI	S	I	V	AN	GO	NGR 232																				
DB	173	RC	V	W	S	D	AL	LT	TT	-A	HT	PA	OV	RD	Y	M	MO	OH	ND	Y	IG	G	I	VA	OR	RV	PT	AD	LL	IG	S	IT	TA	DE	EDK 231																
QY	233	IT	S	D	A	K	M	G	CG	L	L	V	G	G	D	T	V	N	F	L	S	F	S	M	E	F	L	A	K	S	P	E	N	H	O	E	L	I	E	R	P	T	I	P	A	C	E	L	L	R	F 292
DB	232	L	T	E	G	L	V	L	A	E	R	I	L	A	G	E	T	S	A	S	Q	I	P	N	L	V	L	F	R	H	P	O	L	L	E	R	I	R	N	D	H	D	L	P	D	V	E	L	L	R	V 291
QY	293	SL	-V	A	D	G	--R	L	T	S	D	E	R	H	G	V	O	L	K	K	G	D	O	I	L	P	O	M	L	S	G	L	D	E	R	K	N	A	C	P	M	H	D	S	R	K	V	S	H	T 349	
DB	292	P	L	G	T	V	D	G	P	P	R	A	T	E	D	V	E	L	G	V	L	R	A	G	E	T	V	P	S	M	G</																				

RESULT#	5			
ID	054302	PRELIMINARY;	PRT:	404 AA.
AC	054302:			
DT	01-NOV-1996	(TREMBLrel. 01, Created)		
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)		
DT	01-NOV-1999	(TREMBLrel. 12, Last annotation update)		
DE	CYTOCHROME P450.			
CN	RAMP.			
OS	Streptomyces hygroscopicus.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
OC	Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-NRRL 5491;			
RX	MEDLINE: 95372374.			
RA	SCHWECKE T., APARICIO J.F., MOLNAR I., KOENIG A., KHAW L.E.,			
RA	HAYDOCK S.F., OLIVANK M., CAFFREY P., CORTES J., LESTER J.B.,			
RA	BOEHM G.A., STAUNTON J., LEADLAY P.F.;			
RT	"The biosynthetic gene cluster for the polyketide immunosuppressant			
RT	rampamyl."			
RL	Proc. Natl. Acad. Sci. U.S.A. 92:7839-7843(1995).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-NRRL 5491.			
RA	APARICIO J.F., MOLNAR I., SCHWECKE T., KOENIG A., HAYDOCK S.F.,			
RA	KOENIG A., STAUNTON J., LEADLAY F., LESTER J.B., BOEHM G.A.,			
RL	Gene 0.0-0(0).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-NRRL 5491.			
RA	APARICIO J.F., MOLNAR I., SCHWECKE T., KOENIG A., HAYDOCK S.F.,			
RA	EE KHAW L., STAUNTON J., LEADLAY F., LESTER J.B., BOEHM G.A.,			
RA	STAUNTON J., LEADLAY P.F.;			
RL	Gene 0.0-0(0).			
DR	EMBL: X86780; CAA60465.1; -.			
DR	HSSP: Q00441; IOXA.			
DR	PFAM: PF00067; p450. 1.			
DR	PRINTS: PR00359; BP450.			
SO	SEQUENCE 404 AA; 45071 MW; 05AB94DF CRC32;			

9;

```

RESULT          6
ID              Q59819
AC              Q59819;
DT              01-NOV-1996 (TREMBLrel. 01, Created)
DR              01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT              01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE              CYTOCHROME P450 (EC 1.14.14.1).
GN              OLEP.
OS              Streptomyces antibioticus.
OC              Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC              Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN              [1]
RP              SEQUENCE FROM N.A.
RX              MEDLINE: 95255619.
RA              RODRIGUEZ A.M., OLANO C., MENDEZ C., HUTCHINSON C.R., SALAS J.A.;
RT              "A cytochrome P450-like gene possibly involved in oleandomycin
RT              biosynthesis by Streptomyces antibioticus.";
RL              FEMS Microbiol. Lett. 127:117-120(1995).
CC              -I- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR              EMBL: L37000; AAA92553.1; -.
DR              HSSP: Q00441; TOXA.
DR              PROSITE: PS000086; CYTOCHROME_P450; 1.
DR              PFMW: PF00067; P450; 1.
DR              PRINTS: PR00359; BP450.
KW              Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.
FT              BINDING 356 356 HEME (BY SIMILARITY).
SQ              SEQUENCE 407 AA: 34957 MW: 04411660 CRC32:

```

Query Match	Similarity	15.8%	Score 34.5	DB 2:	Length 407;
Match 104	Conservative	63;	Mismatches 168;	Indels 35;	Gaps 11.
QY	60	GGHWATIRGQILIREAYEDYRHFSSSEC---	PTIPRENGEAYDPIPT-----	SMDPPEQ	108
DB	50	GTAAVLVTRMSDARIYVLGDSR--PSTAAATDPATPR-----	MPPPPEDGVLAQDPPDH		101
QY	109	KQFRLANQVAMPVYVDKIENRIQELACSLISLRPOG--CNFTEDYAEPPFIRIFMLLA			167
DB	102	TLRLRLVGAFTARRVEEMRPVRSLVYSLDDMAHNGSPADLVBEFLAAPPVAAICELL			161
QY	168	GLPEEDIPHLKLYLTDOM---TRPDGSMTFEAKKALYDIL--	PIIEOROKPGTDAIS		221
DB	162	GYPLDDROLFRFSDAMLSSTR-----LTAELQIRQYQODPMYMDGLVAORAPFEDLLG			217
QY	222	IVANGQVNCRPIITSDPAKRCGLLVGGIDTVVNFLSFSEMERLANSPEHROELIERPERI			281
DB	218	ALALTDNDHDLTKETIYVMGVSLLIAGHEFSVNOITNLVHLLITERKKRESLVADPALV			277
QY	282	PACCELLLRFSFLVNDG---RLTSDYENHGQVLKKGRDILLPOMISGLDEKKNACPMIV			338
DB	278	PAVVEMLKTYTPLVASGSGVRVATEDVELSTVYRAGECVVHFASANDDEVFHADEL			337
QY	339	DFSROKVSHTTPRGSHSLCLIQHLARRELIIVYLKMLTTRIPDPSA--PCAQIQHNSG--IV			396
DB	338	DEHRENRPIIAGCGAHGHCIGAQRLRELQELASLYVRFRPLDLAEPVAGLKKWQGMIL			397
QY	397	SGCVQALPLVW 406			
DB	398	RGLERQIVSW 407			
RESULT	7				
Q32460					
ID	Q32460	PRELIMINARY;	PRT;	411	AA.
AC	Q32460;				
DT	01-JAN-1998	(TREMblrel. 05, Created)			
DT	01-JAN-1998	(TREMblrel. 05, Last sequence update)			
DT	01-NOV-1999	(TREMblrel. 12, Last annotation update)			
DE	ORF 10.				
OS	Actinodadura hibisca.				
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;				

OC Actinomycetales; Streptosporangineae; Thermomonosporaceae;  
OC Actinomadura.  
RN [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=P157-2;  
RX MEDLINE: 97480928.  
RA DAIRI T., HAMANO Y., IGARASHI Y., FURUMAI T., OKI T.;  
RT "Cloning and nucleotide sequence of the putative polyketide synthase  
RT genes for pradmiclin biosynthesis from Actinomadura hidisaica".  
RL Biosci. Biotechnol. Biochem. 61:1445-1453(1997).  
DR EMBL: D87924; BAA3153.1; -.  
DR HSSP: Q00441; IOXA.  
DR PFAM: PF00067; p450. 1.  
DR PRINTS: PR00359; BP450.  
SO SEQUENCE 411 AA; 44860 MW; 98A620B6 CRC32;

```

Query Match: 15.1%; Score 329; DB 2; Length 411;
Best Local Similarity: 26.4%; Pred. No. 6,4e-19;
Matches 114; Conservative 61; Mismatches 199; Indels 58; Gaps

QY      4 TTGSNANLAPLPVHPVEHLVEFDMDYNP-----SNLSAGVOEAMAVIOESNVPDLVMTGCRN 59
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      9 TVDPPEDVTPAFPPPPD-----DFFQPCCEHARLRRASDPYAKVLP-----T 50
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      60 GGH-MIATRGQLIREAVDYDRHFSSCC---PFITRAGEAEDFTPTS-----MDPE 107
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      51 GDHAWVYTRYADVFRTVSD--RRFSKEAVTRGADR-----LTPMGRSKSLYIMDPE 102
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      108 QROFRLANOVAVGHPVVDKLEENRIIOELACSLIESELRPG--OCNFTEDYAEFPPIRIPML 166
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      103 HTRMKKIVSRFAFTRARVCGMAHVRDLTSGFVDEWMEHGPPADLIARLPLPTVICEM 162
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      167 AGLPEDIDPLKLYLTDQTRDGSMTPEAKE-----ALDYILPIITEORQKPGTDAI 220
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      163 LGVPREDPRKQDMDTRNL-TIGAPALAQADEIKAAVGRGLGYLAELIDAKTAAPADDL 221
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      221 SIYVANGQVNGRPITISDEAKRRCGILLVGLDTPVNFLSFSMEFLAKSPEHROELIERPER 280
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      222 SLSLSAHAMD--GLSEELLTGTGMLLAAGYTTTAALTSHYHLLRPSYARLRREDPSG 280
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      281 IPACEELLRRFSLVADG----RLTSDYEFHGVLKKGQDILLPOMLSGLDEKKNACPM 336
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      281 IPAAEELL-RYGOIGGAGAIRIARVDEVEGTLVRAGEAVIPLFLFNAARDEPEVFADPE 339
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      337 HYDESROKVSHTTGFHGSHLIDGQHLARREIIVTLKEMLRIPDFISA-PGAOTQHKSGI 395
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      340 ELDCRTGNPHIALGHGHGHTGCLAPARLEQVYLFVLVETPRALRLAIDDAIDITWRPGL 399
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      396 V-SGVQALPLVW 406
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      400 AFARPDALPIAM 411
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT      8
059723      PRELIMINARY;      PRT;      406 AA.
AC      AC      059723;
AT      01-NOV-1996 (TREMBLrel, 01, Created)
DT      01-NOV-1996 (TREMBLrel, 01, Last sequence update)
RL      01-NOV-1999 (TREMBLrel, 12, last annotation update)
DE      CYTOCHROME P450 LIN (EC 1.14.14.1).
LN      CYTOCHROME P450 LIN (EC 1.14.14.1).
OS      Pseudomonas incognita.
OC      Bacteria; Proteobacteria.
RN      (1)
RP      SEQUENCE FROM N.A.
RA      ROEP J.D., GUNSALUS I.C., SLIGAR S.G.;
RL      Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
CC      EMBL; L23110; AAA25810.1; -.
DR      HSSP; P33006; ICPT.
DR      PROSITE; PS00086; CYTOCHROME_P450; 1.

```





Db 240 PDEFDPBPSRHIAFAVSHFCGALALRLATVLTLSAISARPPQOLA-GEIYKPNV 298  
QY 395 IVSGVALPL 404  
Db 299 AMRGMALPV 308

## RESULT 13

087192 PRELIMINARY; PRT; 406 AA.

AC 087192; Q59915;  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
DE OEF-2 PROTEIN.  
OS Streptomyces griseus.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA UEDA K.;  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 95378086.  
RA UEDA K., KIM K.M., BEPPU T., HORINOCHI S.;  
RT "Overexpression of a gene cluster encoding a chalcone synthase-like  
RT protein confers redbrown pigment production in Streptomyces griseus.";  
RL J. Antibiot. 48:638-646(1995).  
DR EMBL; AB018074; BAA33494.1; -  
DR HSSP; Q00441; IOXA.  
DR PRAM; PF00067; P450.1.  
DR PRINTS; PR00359; BP450.  
SQ SEQUENCE 406 AA; 45739 MW; 950450FE CRC32;

Query Match 14.5%; Score 317; DB 2; Length 406;

Best Local Similarity 27.7%; Pred. No. 5.9e-18; Matches 106; Conservative 47; Mismatches 202; Indels 28; Gaps 7;

QY 44 VLGESNPDLVWTRNGCHWATRGOLIREAVEDYRHFSSCEPIPREAGEAYDPIPT- 102  
Db 32 LITREPYSRIMAYGEGAMLVTR-----YEDVRYVTTRDRF-SNSAVLGRFPPT- 83  
QY 103 -----MDPEGRQFALANOVVGMVYVKLENRIOLACSLIESLRPGQ-CNF 150  
Db 84 EPIVQAESINIMDPASSRLGLVAKSFTPRRVQMRGTOVVDRDLDEMEEGSPDF 143  
QY 151 TEDYAEFPPIRIFMLAGLPREDIPHLK--YLTDQMTRPDGSMTFAEAKKALYDLPIT 208  
Db 144 VARVASALPLTTTICELDIPADRPWLRAHAMTMNNVGAAGKQDAVRKALRGTFDELT 203  
QY 209 EQRROKPGTDAISIVANGOVNGRPITSDEAKRMCGLLVGLDVTVNFILSFSMEFLAKSP 268  
Db 204 ADDRSPGEDLITSLARQDDELDDDELAVMAMVLLITGQTTTQGLNIATVLTLLTRP 263  
QY 269 EHRBELIERPRIPAAEBELIRFSL--VADGRLTSDYEFHGVALKGGQIILPOMLS 325  
Db 264 DLLSLRAEPPRLPRTLEELRLHLPFRKGVGIPRIALEDVLSGLKAGDVVHVSYLTA 323  
QY 326 GLDERKNACPMHVPDSKOVSHTEFGSHCLGSHLARRELIYTLKMWLRIP--DPSI 383  
Db 324 NRDSAKEDRPDELDPDRPTIPHTFTFGAHHCGLAPLATMELEVAFTLLTRPALRLDV 383  
QY 384 APGAQIOHKSGIVSGVALPLW 406  
Db 384 PPEVSNWNTSTIMRYPLALPV 406

RESULT 14  
Q9X9P7 PRELIMINARY; PRT; 410 AA.

AC Q9X9P7;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
DE NIKF PROTEIN.  
OS Streptomyces tendae.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA BRUNTER C., LAUER B., SCHWARZ W., MOERLE V., BORMANN C.;  
RT "Molecular characterization of co-transcribed genes from Streptomyces  
RT tendae Tug901 involved in the biosynthesis of the peptidyl moiety of  
RT the peptidyl nucleoside antibiotic nikkomycin.";  
RL Mol. Gen. Genet. 0:0-0(0).  
DR EMBL; Y18574; CAB46536.1; -  
SQ SEQUENCE 410 AA; 45884 MW; 9B52CB74 CRC32;

Query Match 14.5%; Score 316; DB 2; Length 410;  
Best Local Similarity 27.1%; Pred. No. 7.2e-18; Matches 98; Conservative 64; Mismatches 180; Indels 20; Gaps 9;

QY 63 WIATRGOLIREAVEDYRHFSSCE-PIIP-----REAGEAYDPIPT-----SMDPEQR 109  
Db 47 WLVLKHDARKLADPRVSADRLHAPPGRLTAQRATTEVRRLSTRSMIHDGDEHG 106  
QY 110 QFRALANOVGMVVDKLENRIOLA-CSLIESLRPGQCNFTEDYAEFPPIRIFMLAG 168  
Db 107 AHRKILGSEFSLRIALRPVQELVDKSIDEMLAAPPALVHEVSGAVPSLVICELLG 166  
QY 169 LPREDIPHL-KYLDQMTRPDGSMTFAEAKALYDLPIT-EQRROKPGTDAISIVANG 226  
Db 167 VPHEQRDRFHEMAGLVSRVSIRERAASDALNDFLEDYTERKERGEPDGLIGRLIAR 226  
QY 227 QVNRPIITSDAKRMCGLLVGLDVTVNFILSFSMEFLAKSPEHQELIERPERIPACE 286  
Db 227 NRRTPVMTDHEIVGTAVMLTAGHQTANMTSLGVALLLENPERIKARLADPSLLPATE 286  
QY 287 ELIRRFSLV--ADGRILTSDYEFHGVALKGGQIILPOMLSGLDERKNACPMHVPDSK 344  
Db 287 EMLTFYSVENAPARVATEDIEIGVTIRKDEGIVSGSLADNMDEVHEHRDRDIEGCA 346  
QY 345 VSHTEFGHSLCGLHARRELIYTLKEMLTIRIPDSIA-PGAQIOHKSGI-VSGVAL 402  
Db 347 RHHVAFGVGHQICGONLARVELEIVETLLRVPGLSLAVPAEELPYKDDAGIYIVRV 406  
QY 403 PL 404  
Db 407 PV 408

RESULT 15  
059831 PRELIMINARY; PRT; 410 AA.

AC 059831;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
DE CYNOCHROME P450 SCA-2 (EC 1.14.14.1).  
OS Streptomyces carboxylus.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SANK 62585;  
RX MEDLINE; 96001248.  
RA WATANABE I., NARA F., SERIZAWA N.;  
RT "Cloning, characterization and expression of the gene encoding  
RT cytochrome P-450sca-2 from Streptomyces carboxylus involved in  
RT production of pravastatin, a specific HMG-CoA reductase inhibitor.";



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 12:35:14 ; Search time 76.18 Seconds  
(without alignments)  
128.722 Million cell updates/sec

Title: US-09-246-451-12

Perfect score: 2179

Sequence: 1 TTTETIQSNANLAPLPPHVE.....IVSGVOALPLWDPATTKAV 414

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2126	97.6	405	Y04128	Pseudomonas putida
2	1178	54.1	446	Y04126	Bacterial and mamm
3	370	17.0	336	R60777	Mythamycin IV hyd
4	338	15.5	408	R7867	S. clavuligerus Or
5	336.5	15.4	587	W33274	S. fradiae tylosin
6	336	15.4	406	R13349	Cytochrome enzyme
7	328	15.1	403	R13350	Cytochrome enzyme
8	327	15.0	411	R13350	Cytochrome enzyme
9	323.5	14.8	398	W54389	Actinomadura hibi
10	319.5	14.7	410	W11585	Streptomyces prist
11	315	14.5	412	R51368	Protein containing
12	296.5	13.6	404	R14724	Sequence of the p4
13	196	9.0	398	R47521	6-hydroxylase enco
14	160	7.3	422	W36128	Vitamin D hydroxyl
15	160	7.3	422	W36132	Daunomycin C-14 hy
16	155	7.1	422	W00729	N-terminal modifie
17	148	6.8	533	R15057	Daunorubicin 14-hy
18	144	6.6	518	W67616	Cytochrome P450C25
19	143	6.6	494	R62825	A. nidulans phenyl
20	135	6.2	512	W93216	Human steroid-21-h
21	133	6.1	508	W35711	Human cytochrome P
22	133	6.1	512	R72365	Chrysanthem flavon
23	133	6.1	512	R93172	Human auxillary cy
24	132	6.1	512	R72365	Human cytochrome P
25	132	6.1	512	W00652	Human auxillary cy
26	132	6.1	512	R93173	Cytochrome P450IA1
27	131.5	6.0	1169	R76544	Human cytochrome P
28	131	6.0	524	P70577	Mitochondrial cyto
29	128.5	5.9	523	R59291	Rat hepato-cytochr
30	128.5	5.9	898	R61030	Rat liver cytochro
31	128.5	5.9	898	P61082	Entire coded seque
32	128.5	5.9	899	P61056	Complete translati
33	128.5	5.9	1144	P81334	translation of pla

34	128.5	5.9	1150	1	P81335	Expression prod. o
35	128.5	5.9	1150	1	P81337	Expression prod. o
36	128.5	5.9	1162	1	P81336	Expression prod. o
37	127	5.8	512	1	R72364	Human auxillary cy
38	127	5.8	512	1	R93171	Human cytochrome P
39	125	5.7	519	1	R34881	Chimeric cytochrom
40	124	5.7	516	1	W67617	P. chrysogenum phe
41	122	5.6	507	1	W92994	Murine 1-alpha-hyd
42	122	5.6	507	1	W93491	Mouse 1-alpha-OHs
43	120.5	5.5	516	1	R72360	Human cytochrome P
44	120.5	5.5	516	1	W00183	Cytochrome P450IA2
45	120.5	5.5	516	1	R93167	Human cytochrome P

ALIGNMENTS

RESULT	1	
ID	Y04128	Y04128 standard; Protein; 405 AA.
AC	Y04128:	
DT	11-JUN-1999	(first entry)
DE	Pseudomonas putida cytochrome P450 protein P450.cam	
KW	Bacterial; mammalian; cytochrome P450; chimeric; fusion protein;	
KW	oxidise; hydrocarbon; carbon-hydrogen bond; hydroxylating;	
KW	bioremediation; environmental pollutant.	
OS	Pseudomonas putida.	
PN	W09908812-A1.	
PD	25-FEB-1999.	
PE	17-AUG-1998; U16979.	
PR	20-AUG-1997; US-056754.	
PA	(UYP ) UNIV ROCHESTER.	
PI	Jones JP, Shimoi M;	
DR	WPI: 99-190131/16.	
DR	N-PSDB; X19926.	
PT	New P450 fusion proteins - comprising a portion of a bacterial	
PT	cytochrome P450 protein and a portion of a mammalian cytochrome P450	
PT	protein	
PS	Disclosure: Page 12-13; 51pp; English.	
CC	The present invention describes a fusion proteins comprising a portion	
CC	of a bacterial cytochrome P450 protein and also a portion of a mammalian	
CC	cytochrome P450 protein. The fusion protein can oxidise hydrocarbons or	
CC	any compound having a carbon-hydrogen bond. The fusion protein can be	
CC	used for hydroxylating a compound to be oxidised. It can also be used in	
CC	the bioremediation of an environmental pollutant. Since the fusion	
CC	protein is soluble, it can be subject to structural elucidation by X-ray	
CC	crystallography for designing functional proteins. It can be readily	
CC	expressed in soil bacteria to facilitate bioremediation. The present	
CC	sequence represents Pseudomonas putida cytochrome P450 protein P450.cam	
CC	from the present invention.	
SQ	Sequence 405 AA:	
Query Match		
Best local Similarity 97.6%; Score 2126; DB 1; Length 405;		
Matches 403; Conservative 1; Mismatches 1; Indels 0; Gaps 0;		
QY	10 NLADLPVPHVBEHLVDFDDMKNPSNLSAGVOEAMVLDSENPVDLVWTRCGHIAIRGQ 69	
DB	1 NLADLPVPHVBEHLVDFDDMKNPSNLSAGVOEAMVLDSENPVDLVWTRCGHIAIRGQ 60	
QY	70 LIRAYVDYRHFSECFPIREAGEAYDFITPSMDPEORFRALANOVGMPPVDDLEN 129	
DB	61 LIRAYVDYRHFSECFPIREAGEAYDFITPSMDPEORFRALANOVGMPPVDDLEN 120	
QY	130 RIQELACSLIESLRPOGQCNTEEDYAPFPRIIFMLAGLPEEDIPHLKYLTDQMPDPG 189	
DB	121 RIQELACSLIESLRPOGQCNTEEDYAPFPRIIFMLAGLPEEDIPHLKYLTDQMPDPG 180	
QY	190 SMTPAEKKEALYDLIRITIEORRQKPGCTDAISYANQVGRPTTSDEAKRMGCLLVGG 249	
DB	181 SMTPAEKKEALYDLIRITIEORRQKPGCTDAISYANQVGRPTTSDEAKRMGCLLVGG 240	

Oy	250	LDYVNFNFSEMEFLAKSPPEHROELTEPPEFLIPACCELLRRSLVNDGHLITSDVEFHG	309
Db	241	LDYVNFNFSEMEFLAKSPPEHROELTEPPEFLIPACCELLRRSLVNDGHLITSDVEFHG	300
Oy	310	VOJLKGDDILLPOMLSGIDEXKRNACPMHVDPSROKVSHTTFGHSHLCLQHLARREIY	369
Db	301	VOJLKGDDILLPOMLSGIDEXKRNACPMHVDPSROKVSHTTFGHSHLCLQHLARREIY	360
Oy	370	TLKEMLTRIPDFSIAPGAOIOHKSGIVSGVQALPLVNDPATTAKV	414
Db	361	TLKEMLTRIPDFSIAPGAOIOHKSGIVSGVQALPLVNDPATTAKV	405

RESULT 2  
 Y04126  
 ID Y04126 standard; Protein; 446 AA.  
 AC Y04126;  
 DT 11-JUN-1999 (first entry)  
 DE Bacterial and mammalian chimeric cytochrome P450 protein.  
 KW Bacterial; mammalian; cytochrome P450; chimeric; fusion protein;  
 KW oxidise; hydrocarbon; carbon-hydrogen bond; hydroxylating;  
 KW bioremediation; environmental pollutant.  
 OS Synthetic.  
 PN MO9908812-A1.  
 PD 25-FEB-1999.  
 PF 17-AUG-1998; U16979.  
 PR 20-AUG-1997; US-056754.  
 PA (DURP ) UNIV ROCHESTER.  
 PI Jones JP, Shimoi M;  
 DR WPI: 99-190131/16.  
 DR N-PSDB: X19916.  
 PT New P450 fusion proteins - comprising a portion of a bacterial  
 PT cytochrome P450 protein and a portion of a mammalian cytochrome P450  
 PT protein  
 PS Claim 24; Page 6-8; Sipp; English.  
 CC The present sequence is a fusion proteins comprising a portion of a  
 CC bacterial cytochrome P450 protein and also a portion of a mammalian  
 CC cytochrome P450 protein. The fusion protein can oxidise hydrocarbons or  
 CC any compound having a carbon-hydrogen bond. The fusion protein can be  
 CC used for hydroxylating a compound to be oxidised. It can also be used in  
 CC the bioremediation of an environmental pollutant. Since the fusion  
 CC protein is soluble, it can be subject to structural elucidation by X-ray  
 CC crystallography for designing functional proteins. It can be readily  
 CC expressed in soil bacteria to facilitate bioremediation.  
 SQ Sequence 446 AA;

Query Match	Similarity	54.1%	Score 1178:	DB 1:	Length 446:
Best Local	Similarity	58.1%:	Prod. No. 7.5e-14:		
Matches 253:	Conservative	30:	Mismatches 100:	Indels 54:	Gaps 10:
QY	10	MLAPLPVPHVPEHLVDFPMYNPSNLISAGVOEAMVLOESNVDELVTWTRCNGSHVIATRQ	69		
Db	1	NLAPLPVPHVPEHLVDFPMYNPSNLISAGVOEAMVLOESNVDELVTWTRCNGSHVIATRQ	60		
QY	70	LIREAYEDYRNHSSCFPIPRAGAGAYDFIPISMPPEQRCQPRALANOVYGMPPYDKLEN	129		
Db	61	LIREAYEDYRNHSSCFPIPRAGAGAYDFIPISMPPEQRCQPRALANOVYGMPPYDKLEN	120		
QY	130	RIQELACSLIESLRQGCQNFTEDEYAEFPFIRIFMLALGLPEEDIPHLKYLLDQMTRPDQ	189		
Db	121	RIQELACSLIESLRQGCQNFTEDEYAEFPFIRIFMLALGLPEEDIPHLKYLLDQMTRPDQ	180		
QY	190	SMTRAEAKELDYILPIITEORQKPGT-----DAISIVANGVQNGRP--ITSDEAKRM	241		
Db	181	SMTRAEAKELDYILPIITEORQKPGNPNODFIDCFLTKMEKRNHNPSEPTISLENT	240		
QY	242	CGLLLVGLDIPVNFLETSFMEFLAKSPE--HROELIER-----DELIPA	283		
Db	241	AVDLEFGAGTETTTSTLKRALLLLKHPETAYAGVOEIERIVIGRNNSPCMQDRSHAPYDA	300		
QY	284	ACEELLRRFSLVADG--RIITSDEYEFHGVOALKGOOILL-----PQMLSG-	326		

Db 301 VVHEQQRIDLLPSTSLPAVYICDKLFENYLLPKCTTLLISLSTVLDHNRKFRPEMDPH 360

QY 327 --LDERKACACPMHVDSEKQKSHHTFGHSHLGLGQNHARREIVTLKEMTLRIPEPSIA 384

Db 361 HFLDDGGN-----FKSKRT-PPPSAGKRICVGFALAGMELPFLTSLILQNFNLKSLV 412

QY 385 PQAQIQHKSGIVSGVALP 403

Db 413 DPKNLD--TTPVNGFASVLP 430

RESULT 3  
R60777  
ID R60777 standard. Protein; 396 AA.  
AC R60777;  
DT 21-JUN-1995 (first entry)  
DE Mycinamicin IV hydroxylating protein.  
KW Mycinamicin; hydroxylase; macrolide; antibiotic; Micromonospora;  
KW Micromonospora griseorubida  
OS Micromonospora griseorubida A11725CN3.  
PN J06235853-A.  
PD 13-SEP-1994.  
PE 09-MAR-1993; 047638.  
PR 09-MAR-1993; JP-047638.  
PA (ASAH ) ASAH KASEI KOGYO KK.  
DR WPI: 94-328997/41.  
DR N-PPSD; 073674.  
DR DNA encoding a protein having mycinamicin IV hydroxylating  
PT activity - for prodn. of mycinamicin, a macrolide antibiotic  
PT Claim 1; Page 12-14; 23pp; Japanese.  
PS The amino acid sequence of a protein having mycinamicin IV hydroxylase  
CC (M) activity. The gene encodes a protein of 396 a.a. The DNA was  
CC obtained from the macrolide antibiotic-producing bacterium Micromonospora  
CC griseorubida A11725CN3/PTYS507. The gene was isolated from the plasmid  
CC pTYS507. The protein encoded by this plasmid can be used to produce  
CC mycinamicin IV in PTYS507-deficient Micromonospora strains.  
SQ Sequence 396 AA;

Query Match	17.0%	Score 370;	DB 1;	Length 396;
Best Local Similarity	28.8%;	Pred. No. 4.6e-30;		
Matches 106;	Conservative 57;	Mismatches 167;	Indels 38;	Gaps 8;
QY	63	WIATGQILIREAYEDYRHSSSECPFI-----PRAGEAYDPIPTSMDDPEQROFRA	113	
Db	43	WLVTI-----YEDYRAVLGDGFVRCGFSMTRDERPTRPREVWKGGLSMDPEHSRLRR	95	
QY	114	LANQVVGMPVYVDKLENRIQELIACSLIESLRPOG-CNFTEDEAEFPPIRIFMLGLPPE	172	
Db	96	LTVKAFPTARRAESLRPRAREIHVELVDQAAATGQADALVAMFARQLPVKVICELLGVSFA	155	
QY	173	DIPHLKYLTDQMTFRDGSN-----TFAAKALADYLIPITIBORROKGTDAISIV	223	
Db	156	D-----HDRFTKRSAGFLSTAETVAEEMQAAQAAVAYMDDLDRRKEPTDLYSAL	208	
QY	224	ANGOVNGRPIITSDEKAKRCGGLLVGGIDPTVNVFLSFSMEFLAKSPENHRELLERPELPA	283	
Db	209	VOARQDQDSLSQELLDLAIGLLVAGYESYTTQIQDVFYLLMTLPRELRLQDLDRPELLPS	268	
QY	284	ACEELLRRFSL--VADGRILITSDYEFVGOLKKGDQIILLPOMLSGLDERKNACPMHNVDF	340	
Db	269	AVEELTRWVPLGVGTAFPRYAVEDYTLKGVITIRAGEPYLASTGANNRDQAQPPADRIDV	328	
QY	341	SRQAKSHTEFGHSGILCGOHLAREILVYTLAKEMLTRIPDPSIA-PCGAQIQKSG-IYSG	398	
Db	329	DRTPNQHLGFGHGVHICGAPLARVELQVALEVLRLQRLGIRLGIPIETQLRMSSEGMLLNG	388	
QY	399	VOALPLVN 406		
Db	389	PLELPVW 396		



R77867  
ID R77867 standard; Protein; 408 AA.  
AC R77867;  
DT 13-NOV-1995 (first entry)  
DE S. clavuligerus ORF10 product.  
KW Clavulanic acid; clavulinic acid; antibiotic; beta-lactamase-inhibitor.  
OS Streptomyces clavuligerus.  
PN CA2108113-A.  
PD 09-APR-1995.  
PF 08-OCT-1993; 108113.  
PR 08-OCT-1993; CA-108113.  
PI (UVAL-) UNIV ALBERTA.  
PA Aldoo KA, Jensen SE, Paradkar AS;  
DR WPI: 95-207301/28.  
DR N-PSDB: 091580.  
PT Clavulanic acid biosynthesis enzymes and corresp. DNA - useful for  
PT biosynthesis of the antibiotic in Streptomyces hosts which do not  
PT naturally produce clavulanate  
PS Claim 32; Fig.19; 41pp; English.  
CC A 15 kb fragment S. clavuligerus NRRL 5741 genomic DNA (091580),  
CC extending downstream from pcdc, included 10 ORFs encoding the  
CC enzymes required for clavulanate biosynthesis. The ORF10  
CC product (R77867) showed high similarity to cytochrome P450-type  
CC enzymes from other Streptomyces spp.  
SQ Sequence 408 AA;

Query Match 15.58; Score 338; DB 1; Length 408;  
Best Local Similarity 29.38; Pred. No. 1e-26;  
Matches 123; Conservative 59; Mismatches 184; Indels 54; Gaps 18;

16 PHVEHLEFDEDMYNPSLSAGVOEAMAVLQESNPDLWTRNGGH-WIATRGQLIREA 74  
14 PAYPMHRCVPD---PPQLAGLSQKASRYT-----LW---DGSQVWLVSAGARAV 62  
75 YEDYRHFS-SECPETP-----REAGEAYDEIPTSMDEPEQROFRA-----LANOV 118  
63 LGDRRFATVATPCGPMRLTRTSQLYRANPESAFI--RMDPDQHSRLKSMLTRPFLARRA 120  
119 VGM-PVYDKLENRIQELACSLIESLRPOGQCNFTEDVAEPPPIRIFMLAGLPREDIPHL 177  
121 EALRPVAVEL---LDEILGLVKGSRP---VDLVAGLTIPVSRVITLLFCAGDDRREFI 174  
178 K-----YLTDDQTRPDGSMFAKAEALVDYLPIIEQRORRGGTDAISIVANGOVNRP 233  
175 EDRSAVLIDKRYTPE--QVAKARDELGYRLREVEERIEPGLDLSRLVIDOVRGHL 231  
234 TSDEAKRMCGILLVGLDTVNVNPLFSMEFLAKSPENHROELIERPELIPACEELLRRFS 293  
232 RVEENVPRCRLLIAGHGTTSQASLSLSLTLTPELAGRLTEBPALLPKAVEELLRRFS 291  
294 LVADG--RILTSDEYEFHGVOLKKGQDILLPOMLSGLDERKNACPMHWDSRQKVSHTTEG 351  
292 IVONGLAFAAIVEDQDLDDVLIRAGEGVVLSAGNRDETVPDPDRVVDDBARHHLAFG 351  
352 HGSMLCGIHLAR---REIITYLKEMWLTRIPDSIA--GCAIQHKSGLVS--GYQALPLVM 406  
352 HGMHOCIGOMLARELELLELAIVLRKM---PGARLAVFEELDFRHEVSYSGALPLVTW 408

RESULT 5  
ID M33274 standard; Protein; 587 AA.  
AC W33274;  
DT 17-FEB-1998 (first entry)  
DE S. fradiae tylosin biosynthesis gene product tyIH.  
KW Tylosin; biosynthetic gene product; production; antibiotic; tyIH.  
OS Streptomyces fradiae.  
PN US672497-A.  
PD 30-SEP-1997.  
PF 21-DEC-1995; 575843.  
PR 12-MAY-1989; US-351350.  
PR 21-MAR-1986; US-842330.

PR 25-JUL-1986; US-890670.  
PR 24-FEB-1987; US-018237.  
PR 06-AUG-1991; US-742222.  
PR 28-JUL-1993; US-107232.  
PR 17-FEB-1994; US-196872.  
PR 21-DEC-1995; US-575843.  
PA (ELIL) Lilly & CO ELI.  
PI Cox KL, Fishman SE, Hersberger CL, Seno ET;  
DR WPI: 97-48860/45.  
DR N-PSDB: T58686.  
PT DNA encoding Streptomyces fradiae tylosin biosynthesis gene products  
PT - for increasing tylosin production in Streptomyces spp.  
PS Claim 20; Columns 21-38; 38pp; English.  
CC The present sequence is the Streptomyces fradiae tylosin  
CC biosynthetic gene product tyIH, useful to increase the production  
CC of the antibiotic tylosin in Streptomyces spp.  
SQ Sequence 587 AA;

Query Match 15.48; Score 336.5; DB 1; Length 587;  
Best Local Similarity 29.68; Pred. No. 2.6e-26;  
Matches 106; Conservative 62; Mismatches 169; Indels 21; Gaps 11;

63 WIATRGQLIREAYDYR---HFSECEPTPREAGEA-YDFIPTSMDEPEQRFALANOV 118  
143 WLISRODHVRLALADPRYSIH-PAKLRLSPDGEAEASRSLLTLDPPDHALGHFTPE 201  
119 VGMPEYDKLENRIQELACSLIESLRPOG-QCNFTEDVAEPPPIRIFMLAGLPREDIPHL 177  
202 FGLRRVRELRSRPSQIVYGLGLDILTARGDEADLLADFALPMATQVICHLLDIPYEDRYF 261  
178 KYLDQMTRPDGSMTFAKAEALVDYLPIIEQRORRGGTDAI--SIYANOVNCRPTS 235  
262 QERTQATRPAGEBALLELDLRLDRLISGTRGREGGMLGSMA--QARGGGISH 319  
236 DEAKRMCGILLVGLDTVNVNPLFSMEFLAKSPENHROELIERPELIPACEELLRRFS 295  
320 ADVLDNAVLLLAAGHETTSQASLSLSLTLTPELAGRLTEBPALLPKAVEELLRRFS 378  
296 ADG--RILTSDEYEFHGVOLKKGQDILLPOMLSGLDERKNACPMHWDSRQKVSHTTEG 353  
379 ADGLRSATADIEIDCHIRAGDCILVFLAANRDEAVFSEPAFDIHRARRHVAFCYG 438  
354 SHLCLGHLARLEIIVLTKEMWLTRIPDSIAP-----GCAIQHKSGLVS--GYQALPLVM 406  
439 PHOCIGOMLARELELLELAVGLERLP--ALRPITDVAGLRKSDSA--VFGYELPLVM 493

RESULT 6  
ID R11349 standard; Protein; 406 AA.  
AC R11349;  
DT 05-JUN-1991 (first entry)  
DE Cytochrome enzyme P450SUI.  
KW Cytochrome P450; P450SUI; P450S2; herbicide resistance.  
OS Streptomyces griseolus.  
PN WO9105561-A.  
PD 21-MAR-1991.  
PD 27-AUG-1990; 004785.  
PF 11-SEP-1989; US-405605.  
PR 12-JAN-1990; US-464499.  
PR 23-AUG-1990; US-569781.  
PA (DOPO) DU PONT DE NEMOURS CO.  
PI Dean G, Harder PA, Lelo KJ, Lichner FT, Odell JR;  
PI O'Keefe DP, Omer CA, Romesser JA;  
DR WPI: 91-102077/14.  
DR N-PSDB: 011126.  
PT DNA encoding cytochrome P450 enzymes - and electron donating  
PT iron sulphur proteins, used to confer herbicide resistance to  
PT plants and microorganisms  
PS Claim 13; page 151; 224pp; English.  
CC This cytochrome P450 enzyme, P450SUI is expressed alongside the iron  
CC sulphur protein Fes-B, by a DNA sequence contained in a recombinant



Db 51 GDHAWTVTRIVADVREVTSD--RRFSKAVTRPGADP-----LIPMGRSKSLVIMDPE 102  
 QY 108 QROFRALANQVGMVYVDKLENRIQELACSLIESLRPOG--OCNFTEDYAEPPPIRIFMLL 166  
 Db 103 HTBMKIVYSRAFTARVRGMAHVRDLTSGVDEKVEHGRPADLIALHLALPLPTVICEM 162  
 QY 167 AGLPEDIPHLKYLTLDQKTRPDGSMTPFAKE-----ALYDYLPIIEQRQRKGTDAI 220  
 Db 163 LGVPEDRRFQDWTDRML--TIGAPALAOADEIKAAVGRGLYLAELIDAKTAAPADLL 221  
 QY 221 SIVANGQVNGRPITSDEKRRMCGLLVGGDLTVVNFLSFSEFLAKSEHGOELLEREL 280  
 Db 222 SLTSRAHAD--GLSEELITFGMTLLAAGYHTTAATHSVYHLRPSRYARLRKEDPSG 280  
 QY 281 IPACEELLRRFSVLAVD----RLTSPDEPHGVOLKKGDQILLPQMLSGDERKNACPM 336  
 Db 281 IPAAVEELL--RYGQIGGACAGIRIAVEDYEVGGLIVRGEAVIPLFNANANDPEVFADPE 339  
 QY 337 HVDSEKRVSHTEFGHSHLCLGHLARREIIVTLKEMLTIRIPDSIA--PGAQIOHKSQI 395  
 Db 340 ELDGRFTDNPHIALGHGTHYCLGAPLARLELOVLETLVERTPALRLAIDADITWRPGL 399  
 QY 396 V-SCVQALPLTW 406  
 Db 400 AFARPDALPIAW 411  
 RESULT 9  
 ID W11585 standard; Protein; 398 AA.  
 AC W11585;  
 DT 02-APR-1997 (first entry)  
 DE Streptomyces pristinaespiralis Smbf gene product.  
 KW Streptogramin B: antibiotic; biosynthesis; pristinamycin;  
 KM virginiamycin; dipecolic acid; cyclodeamination; papa; snba; snbf;  
 OS Streptomyces pristinaespiralis.  
 PN MO9601901-A1.  
 PD 25-JAN-1996.  
 PE 04-JUL-1995: F00889.  
 PR 08-JUL-1994: FR-008478.  
 PA (RHON ) RHONE POULENC RORER SA.  
 PI Barriere JC, Blanc V, Crouzet J;  
 PI Debussche L, Paris JM, Thibaut D, Bamas-Jacques N;  
 PI Dutruc-Rosset G, Famechon A;  
 DR WPI: 96-097631/10.  
 DR N-PSDB: T58555.  
 PT New streptogramin B derivs. useful as antibiotics - produced by new  
 PT mutants of Streptomyces having altered genes for streptogramin B  
 PT biosynthesis  
 PS Example 1: Page 113-114; 146pp: French.  
 CC The papa gene of S.pristinaespiralis is involved in the biosynthesis  
 CC of 4-dimethylamino-L-phenylalanine (DMPAPA), a precursor for  
 CC pristinamycin 1A. Upstream of the papa gene, on the complementary  
 CC strand, is the snba gene coding for 3-hydroxypicolinic acid-AMP ligase.  
 CC The region between these two genes was sequenced and two open reading  
 CC frames were identified. The first (pipa) decodes to an amino acid  
 CC sequence with homology to ornithine cyclodeaminase from Agrobacterium  
 CC tumefaciens. The pipa gene product is likely to catalyse the cyclo-  
 CC demethylation of lysine, leading to production of dipecolic acid.  
 CC Mutations in the pipa gene were shown to affect dipecolic acid  
 CC synthesis but not the synthesis of 3-hydroxypicolinic acid. The second  
 CC open reading frame (snbf) could be decoded to give a product with  
 CC homology to hydroxylases of the cytochrome P450 type.  
 CC Disruption of the pipa and snbf genes can be used to produce  
 CC strains of S.pristinaespiralis which are unable to produce the  
 CC antibiotic pristinamycin I but which may be able to produce new,  
 CC modified forms of it.  
 SQ Sequence 398 AA;  
 Query Match 14.8%; Score 323.5; DB 1; Length 398;  
 Best Local Similarity 29.0%; Pred. No. 3.1e-25;

Matches 106; Conservative 59; Mismatches 159; Indels 41; Gaps 12;  
 QY 74 AVEYRH-----FSSECPFIPREAGEAVDPIPTSMDEPEORQFRALANQVGM 121  
 Db 36 AFHYFRHADVLTVASDPRGVYSQSLRRLRGQALSSEQLSYIDPPMHTLRLVLSQAFTP 95  
 QY 122 PVDKLENRIQELACSLIESLRPOG--NFTEDYAEPPPIRIFMLLAGLPEDIPHLKYLT 180  
 Db 96 RTVADLEPRVETLAGQLDAV--DQDFDLVADFAVPLPVIVIAELLGVPRADPTLFRSM 153  
 QY 181 TDQMT-----PGSMTFAKAKALDYLIPIIEQRQRKGTDAISIVA 224  
 Db 154 SDRMLQMVADPADMQFDDADEDYQRLVKPEPMAMAHYLIHDHTDRARPANDLISALV 213  
 QY 225 NGVNGRPITSDEKRRMCGLLVGGDLTVVNFLSFSEFLAKSEHGOELLEREL--LIP 282  
 Db 214 AARBERLTDEQIYERGCALLLMAGHVSTSMILGNVYCLDHP--RKAARADRSLTP 271  
 QY 283 AACCELLR-RFSVLAVDGRILTSDEPHGVOLKKGDQILLPQMLS--GLDERKNACPMHYDF 340  
 Db 272 ALIEVELRLRPITVYMARVTTKQDVLAGTTPAG--RMVVPSELNANDEQVFTPDHIDL 330  
 QY 341 SRQVSHTEFGHSHLCLGHLARREIIVTLKEMLTIRIPDSIAPGAQIO--HKSQIVSGV 399  
 Db 331 AREG-ROIAFGHGHYICGAPLARLEGRILAEALFDRPDESPDGAKLVYHROGLF--GV 388  
 QY 400 QALPL 404  
 Db 389 KNLPL 393  
 RESULT 10  
 ID R51368 standard; Protein; 410 AA.  
 AC R51368;  
 DT 24-NOV-1994 (first entry)  
 DE Protein containing Cytochrome P450 SCA-2 activity.  
 KW Cytochrome P450 SCA-2; Streptomyces carbophilus;  
 KM treatment of hyperlipidaemia; drug preparation.  
 OS Streptomyces carbophilus.  
 FH Key Location/Qualifiers  
 FT protein 1..410  
 FT J06070780-A.  
 FN 15-MAR-1994.  
 PD 28-AUG-1992; 229969.  
 PR 28-AUG-1992: JP-229969.  
 PA (SANY ) SANKYO CO LTD.  
 DR WPI: 94-128679/16.  
 DR N-PSDB: 061452.  
 PT Cytochrome P450 SCA-2 gene - from Streptomyces carbophilus,  
 PT useful in treatment of hyperlipidaemia  
 PS Claim 1; Page 12-14; 18pp: Japanese.  
 CC R51368 shows a protein having cytochrome P-450 activity. P-450  
 CC SCA-2 can be prepared commercially for use in the preparation of a  
 CC drug for the treatment of hyperlipidaemia.  
 SQ Sequence 410 AA;  
 Query Match 14.7%; Score 319.5; DB 1; Length 410;  
 Best Local Similarity 25.9%; Pred. No. 8.5e-25;  
 Matches 93; Conservative 63; Mismatches 184; Indels 19; Gaps 7;  
 QY 63 WIATRGQILREAYEDYR-----HSSSECPFIP--REAGEAVDPIPTSMDEPEORQFR 112  
 Db 56 WVVYTKHEAARLLADPRLSRDRLHADFPATSPRKAFFQSGSPAF----IGMDPEHGTRR 111  
 QY 113 ALANQVGMVYVDKLENRIQELACSLIESLRPOG--OCNFTEDYAEPPPIRIFMLLAGPE 171  
 Db 112 RMTISEFTVKRIKGRPDVERIVHGFTDDMLAAGPVALVSOFALPVYSWVICMLGVPY 171  
 QY 172 EDIPHLKYLTQDMTRPDGSMTPFAKAKALDYLIPIIEQRQRKGTDAISIVANGQVNGR 231

```
Dh 172 ADHEFDODASRLVQAVDADASAAARDFFERYLDGLITLKESRPGTGLGLVTHQADG 231
Qy 232 PITSDEAKRMCGLLVGGIDTVVNFSLFSMEFLAKSPENROELIERPELLPAACEELLR 291
Dh 232 EIDRAELISTALLLVGHSHETASMTSLVTTLEHPDQHALLRADPSLVGAVPELLRV 291
Qy 292 FSL--VADGRILTSYERFHVQVKKDQOILLPQMLSGIDERKNACPMHVDERSKQVSHTT 349
Dh 292 LAIDINGGRATATADIEIDGOLIRAGEVITVNSIANRDSVFPENPRDLVHRSAHHLS 351
Qy 350 FGHSHLICGHLARREIYVTLKEMLTRIPDFSTA-PGAQIQHKS-G-IVSGVQALLPLVW 406
Dh 352 FGTVGHQICGHLARLEVLITVLPDRITPLRLAVPEVQLLTRPGTTIQGVNELPVTW 410

RESULT 11
ID R38309 standard; Protein; 412 AA.
AC R38309;
DT 04-DEC-1993 (first entry)
DE Sequence of the P450-soy protein.
KW P450soy; soy; gene; oxidation; haem protein.
OS Streptomyces griseus ATCC 13273.
PN M09312235-A.
PD 24-JUN-1993.
PF 16-DEC-1992; U10885.
PI 16-DEC-1991; US-807001.
PA (DUPO ) DU PONT DE NEMOURS & CO E. I.
PI Omer CA, Sariaslani FS, Trower MK;
DR WPI: 93-214178/26.
DR N-PSDB; Q45569.
PT Constitutive expression of P450 SOY (SOYC) and ferredoxin soy
PS (soy) in Streptomyces - used for oxidn. of organic chemicals
PS Example; Figure 2: 45pp; English.
CC Cytochrome P450soy was purified from S. griseus ATCC 13273. Two
CC similar forms of P450soy were isolated. P450soy-delta, is derived
CC from P450soy by in vitro proteolysis during isolation. One of the
CC tryptic peptide fragments of cytochrome P450soy and of of the
CC P450soy-delta protein were subjected to automated degradation. The
CC NH2 terminal sequences are given in R38306 and R38307. A mixture of
CC oligos that consist of possible DNA sequences that could encode the
CC AAS FGYNHCL of the tryptic peptide was made. It consists of the
CC sequences in Q43290-23. The oligo mixture was end-labeled and used
CC to probe the EMBL4 library of S. griseus DNA. Hybridizing plaques
CC were isolated and a 4.8kb SacI DNA fragment was isolated from one
CC clone that hybridised to the oligo probe mixture. As segment of the
CC 4.8kb fragment was sequenced and found to contain an ORF. Within
CC this ORF was a section that corresp. exactly to the AA sequence
CC determined from the cytochrome P450soy tryptic peptide (see Q45569,
CC R38309). The gene encoding the P450soy protein was called soyC. Five
CC nucleotides downstream from the stop codon for soyC another ORF was
CC identified. This ORF encodes an apparent ferredoxin-like protein.
CC The gene was designated soyb and the protein ferredoxin-soy.
SQ Sequence 412 AA;

Query Match 14.5%; Score 315; DB 1; Length 412;
Best Local Similarity 27.2%; Pred. No. 2,5e-24;
Matches 84; Conservative 56; Mismatches 163; Indels 6; Gaps 5;
```

```
Dh 285 AVVVEELL-RFLSLAEGLQRLATFEDMEVDATIRKGGVFNSTSLINRDADVPERAETLD 343
Qy 340 FSRQKVSHTTFGHSHLICGHLARREIYVTLKEMLTRIPDFSTA-PGAQIQHKS-G-IVS 397
Dh 344 WDRPARHHLLAFGFGVHCOGLNARAEILDAMRTYLERPLQRLAVPAHEIRHKPGDTIQ 403
Qy 398 GVOALPLVW 406
Dh 404 GLDPLPVAM 412

RESULT 12
ID R14724 standard; Protein; 404 AA.
AC R14724;
DT 28-JAN-1992 (first entry)
DE 6-hydroxylase encoded by eryF gene.
KW C-6 hydroxylation; erythromycin; 6-deoxyerythromycin; antibiotics;
KW saccharopolyspora; cytochrome P450 monooxygenase; ss.
OS Saccharopolyspora erythraea.
PN M09116334-A.
PD 31-OCT-1991.
PF 16-APR-1991; U02600.
PI 18-APR-1990; US-510483.
PA (ABBO ) ABBOTT LABORATORIES.
PI Weber JM;
DR WPI: 91-339744/46.
DR N-PSDB; Q14548.
PT New 6-deoxyerythromycin derivs. - are antibiotics with increased
PT acid stability, produced by cultivation of saccharopolyspora.
PS disclosure; Fig 3; 56pp; English.
CC The eryF gene encodes the 6-hydroxylase component of the cytochrome
CC P450 monooxygenase system responsible for the hydroxylation of 6-de-
CC oxyerythronolide B to erythronolide B. Interruption of this step
CC results in the formation of deoxyerythromycin A and new derivatives
CC useful as antibiotic which have better stability against acids that
CC the corresponding erythromycins. Interruption of the reaction can
CC be effected by an insertion into the eryF gene of a plasmid, gene
CC replacement or chemical or light-induced mutagenesis.
CC See also Q14549.
SQ Sequence 404 AA;

Query Match 13.6%; Score 296.5; DB 1; Length 404;
Best Local Similarity 24.2%; Pred. No. 2e-22;
Matches 90; Conservative 74; Mismatches 173; Indels 35; Gaps 10;
```





---

**THIS PAGE BLANK (USPTO)**









SEQ ID NO 8  
LENGTH: 419  
TYPE: PRF  
ORGANISM: Sorangium cellulosum  
US-09-335-409-8

Query Match 12.7%; Score 277.5; DB 3; Length 419;  
Best Local Similarity 24.0%; Pred. No. 9, 8e-21;  
Matches 101; Conservative 67; Mismatches 178; Indels 75; Gaps 14;

QY 6 QSNANLAPLPHVPEHIVPEFDMYNSNSAGVQ-----AMAVLOSNNVDLVWTRCNG 61  
DB 3 QEQANOSERKP-----AFDFKPPAP-----GYAEDFPFAIERLRRA-TPIFYWD--EGR 48  
QY 62 HWIATRGQ-----LIREAYEDYHFSSECFIRREAGEAVDFPTSMDDPEQR 109  
DB 49 SWVUTRIHDVSAVFRDERFAVSRREWSSAEYSSAIP-----ELSDKKYGLFGLPEPDHA 104  
QY 110 QFRALANQVGMFVVDKLENRIQELACSLIESLRPOQCNETDYAEPPRIRIFMLLAGL 169  
DB 105 RVKRLVNPSTSRALDILRAEIQRTVDQLDARSQGEFFVDVDAEGIPRAISALTKV 164  
QY 170 PEEDPHILKYLTDQMTRPDGSMTFAEKKALYLLP-----I 207  
DB 165 PAF-----CDEKRRRFSAT-----ARALGVLPVQVDETKTLVASVTGLALHDV 212  
QY 208 IEORROKP-GTDAISIVANGVGRPTTSDCAKRMGGLLVGGLDVVNFPSMEFLAK 266  
DB 213 LDRRRNPLENDVLTMLQAEADGSRSTELVALYCAITADTTIYIAAVALNLR 272  
QY 267 SPEHROELIERPELLPACCELLRPSLVADG--RILTSDYEFHGVOLKKGOI--LLPQ 322  
DB 273 SPEALELVKAEPLMRNALDEVLFNDILRIGTVRARODLEYCGASIKKGEWVFLIPS 332  
QY 323 MSLGLDERKACPMHVDFSKQVSHTFEGSHLCLGCHLAREIIVTLKWLTRIPDFS 362  
DB 333 ALR--GCTVFSRDPVDVDRDGTASLACGPHVCPSGLARLEAELAVGTIRREPDK 390  
QY 383 I 383  
DB 391 L 391

RESULT 7  
US-09-096-982-5  
Sequence 5, Application US/09096982  
Patent No. 5962293  
GENERAL INFORMATION:  
APPLICANT: Strohl, William R.  
APPLICANT: Dickens, Michael L.  
APPLICANT: Desanti, Charles L.  
TITLE OF INVENTION: METHOD OF PRODUCING DOXORUBICIN  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CALFEI, HALTER & GRISWOLD  
STREET: 800 Superior Avenue, Suite 1400  
CITY: Cleveland  
STATE: Ohio  
COUNTRY: USA  
ZIP: 44114-2688  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/096,982  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Goltick, Mary E.  
REGISTRATION NUMBER: 34829

REFERENCE/DOCKET NUMBER: 22727/00131  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 216-622-8458  
TELEFAX: 216-241-0816  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 422 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-096-982-5

Query Match 7.3%; Score 160; DB 2; Length 422;  
Best Local Similarity 23.6%; Pred. No. 2, 1e-08;  
Matches 96; Conservative 51; Mismatches 205; Indels 54; Gaps 14;

QY 39 QFAMAVLOSNNVDLVWTRCNGHWIATRGQOLIREAYEDYRHFSSE--CPFIPREAGEAY 96  
DB 31 REAGPVV-EVNAPE-----AGPRAWITDDALAREVLADPRFVKGPDLPATAWRGVDDL 83  
QY 97 D-----FPTSMDDPEQRORALANQVGMFVVDKLENRIQELACSLIESL-----R 143  
DB 84 DIPVPELRPTTLTAVDGEHRRRLRIHAPFNPRRLAERTDRIATAIDRLTELADSSDR 143  
QY 144 POQCNETDYAEPPRIRIFMLLAGLP-----EEDIPHILKYL-----TDQMTRPD 188  
DB 144 SGPRALIGGFANHFPLVITCELLGVPRVDPMARAEVGLKAGQPSAGGSDTDR 203  
QY 189 GSMTEAKAEALYDYLPIIEORROKPGTDAISIVANGVGRPTTSDCAKRMGGLLVG 248  
DB 204 GDVPRVSALESL--LLEAVHAARKDPTMTBTVLYERRADEFVSDDDLVWYITGLIPA 261  
QY 249 GUDTVVNFPSMEFLAKSPEHROELIERPELLPACCELLRR-----PSYVADGRIILS 303  
DB 262 GUDTGSFLGF--LAEVLARLADADGDAISRVEALRHHPVPSL--WRFPAAT 315  
QY 304 DYEFHGVOLKKGOIILLPMLSGDERKKNACPMHVDFSKQVSHTFEGSHLCLGCHLA 363  
DB 316 EYVIRGVRLPRGAPVLVDLEGTTDGRHHHDARHAFPRDRSRRLRTFGGPHNICGEOLA 375  
QY 364 RREIIVTLKWLTRIPDFSIA--PGAQIQ--HNSGIVGVQALPLVW 406  
DB 376 QLESFTMIGVLRSRFPQARLAVYEELRWCKRGAQYARLTDLR-VW 420

RESULT 8  
US-08-653-650A-5  
Sequence 5, Application US/08653650A  
Patent No. 5976830  
GENERAL INFORMATION:  
APPLICANT: Strohl, William R.  
APPLICANT: Dickens, Michael L.  
APPLICANT: Desanti, Charles L.  
TITLE OF INVENTION: METHOD OF PRODUCING DOXORUBICIN  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CALFEI, HALTER & GRISWOLD  
STREET: 800 Superior Avenue, Suite 1400  
CITY: Cleveland  
STATE: Ohio  
COUNTRY: USA  
ZIP: 44114-2688  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/653,650A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:



```

; APPLICATION NUMBER: US/08/653,650A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldick, Mary E.
; REGISTRATION NUMBER: 34829
; REFERENCE/DOCKET NUMBER: 22727/00131
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 216-622-8458
; TELEFAX: 216-622-8458
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 474 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-653-650A-8

Query Match 7.3%; Score 160; DB 2; Length 474;
Best Local Similarity 23.6%; Pred. No. 2.5e-08;
Matches 96; Conservative 51; Mismatches 205; Indels 54; Gaps 14;

QY 39 QEAMAVLQESNVDPDLVWTRCGNGHMITRGQLIREAYEDYRHSS--CPPIPREAGEAY 96
DB 83 REAGPVV-EVNAP-----AGPRAWITTDALAREVLADRFVKGPDLPATMAGVDDGL 135
QY 97 D-----FPTSMDEPQROFRALANQVVGMPVVDKLENIQELACSLIESL-----R 143
DB 136 DIPVELRPPTLLAVDQEDHRLRIHAPAFNPRLAERTDRIATAIDRLTELAOSSDR 195
QY 144 PQGCNCTEDYAEPPRIRITMLAGLP-----EEDIPMLKYL-----TDQMPRPD 188
DB 196 SGEPAELIGGFAYHFLPLVICELGVPVTPDPMAREAVGVKALGLGPGSAGSDGTDP 255
QY 189 GSMFAFAKALDYLLPIEDOROKPGTDAISVANGOVNGRPTTDEAKRMCGLLVG 248
DB 256 GVPDPTSALES--LEAVNAARRKDRTMTRVLYERAQAEFGSVSDOLVYMTGLIFA 313
QY 249 GLDVTNVFLFSMEFLAKSPENROELIERPELLPAACEELLRR-----FSLVADGRILTS 303
DB 314 GHDTGSFLGF--LLAEVLAGRLADADGDALSRFVEALRHHPVRYTSL--WRFAAT 367
QY 304 DYEHGVQAKKGQDQILLPQMLSGDERKNACPMHNDPSROKVSHTTFGHSILCQHLA 363
DB 368 EYVLRGVRLPRGAPVLVDIGTNDGRHNDAPHAHFHPRDRSRRLTFGDGRHCTGSQLA 427
QY 364 RREIIVLKEMLRIRIPFESIA-PGAQIO--HKSGIVSGVALPLVW 406
DB 428 QLESRTMIGVLRSPQARLAVPYEELRWCRKGQTAARLTDLR-VW 472

RESULT 11
US-09-096-982-9
; Sequence 9, Application US/09096982
; Patent No. 5962293
; GENERAL INFORMATION:
; APPLICANT: Strohl, William R.
; APPLICANT: Dickens, Michael L.
; APPLICANT: Desanti, Charles L.
; TITLE OF INVENTION: METHOD OF PRODUCING DOXORUBICIN
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CALFEE, HALTER & GRISWOLD
; STREET: 800 Superior Avenue, Suite 1400
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: USA
; ZIP: 44114-2688
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```

; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/096,982
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldick, Mary E.
; REGISTRATION NUMBER: 34829
; REFERENCE/DOCKET NUMBER: 22727/00131
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 216-622-8458
; TELEFAX: 216-241-0816
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 443 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-096-982-9

Query Match 7.2%; Score 157; DB 2; Length 443;
Best Local Similarity 23.6%; Pred. No. 4.6e-08;
Matches 95; Conservative 50; Mismatches 210; Indels 48; Gaps 13;

QY 39 QEAMAVLQESNVDPDLVWTRCGNGHMITRGQLIREAYEDYRHSS--CPPIPREAGEAY 96
DB 52 REAGPVV-EVNAP-----AGPRAWITTDALAREVLADRFVKGPDLPATMAGVDDGL 104
QY 97 D-----FPTSMDEPQROFRALANQVVGMPVVDKLENIQELACSLIESL-----R 143
DB 105 DIPVELRPPTLLAVDQEDHRLRIHAPAFNPRLAERTDRIATAIDRLTELAOSSDR 164
QY 144 PQGCNCTEDYAEPPRIRITMLAGLP-----EEDIPMLKYL-----TDQMPRPD 188
DB 165 SGEPAELIGGFAYHFLPLVICELGVPVTPDPMAREAVGVKALGLGPGSAGSDGTDP 224
QY 189 GSMFAFAKALDYLLPIEDOROKPGTDAISVANGOVNGRPTTDEAKRMCGLLVG 248
DB 225 GVPDPTSALES--LEAVNAARRKDRTMTRVLYERAQAEFGSVSDOLVYMTGLIFA 282
QY 249 GLDVTNVFLFSMEFLAKSPENROELIERPELLPAACEELLRRFSV--ADGRILTS 306
DB 283 GHDTGSFLGF--LLAEVLAGRLADADGDALSRFVEALRHHPVRYTSL--WRFAAT 339
QY 307 FHGVQAKKGQDQILLPQMLSGDERKNACPMHNDPSROKVSHTTFGHSILCQHLAR 366
DB 340 IRGVRLPRGAPVLVDIGTNDGRHNDAPHAHFHPRDRSRRLTFGDGRHCTGSQLA 399
QY 367 IIVTLKEMLRIRIPFESIA-PGAQIO--HKSGIVSGVALPLVW 406
DB 400 SRTMIGVLRSPQARLAVPYEELRWCRKGQTAARLTDLR-VW 441

RESULT 12
US-08-653-650A-9
; Sequence 9, Application US/08653650A
; Patent No. 5976830
; GENERAL INFORMATION:
; APPLICANT: Strohl, William R.
; APPLICANT: Dickens, Michael L.
; APPLICANT: Desanti, Charles L.
; TITLE OF INVENTION: METHOD OF PRODUCING DOXORUBICIN
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CALFEE, HALTER & GRISWOLD
; STREET: 800 Superior Avenue, Suite 1400
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: USA
; ZIP: 44114-2688
; COMPUTER READABLE FORM:
```

```

1 MEDIUM TYPE: Floppy disk
2 COMPUTER: IBM PC compatible
3 OPERATING SYSTEM: PC-DOS/MS-DOS
4 SOFTWARE: Patentin Release #1.0, Version #1.30
5 CURRENT APPLICATION DATA:
6 APPLICATION NUMBER: US/08/653,650A
7 FILING DATE:
8 CLASSIFICATION: 435
9 ATTORNEY/AGENT INFORMATION:
10 NAME: GOLPICK, Mary E.
11 REGISTRATION NUMBER: 34829
12 REFERENCE/DOCKET NUMBER: 22727/00131
13 TELECOMMUNICATION INFORMATION:
14 TELEPHONE: 216-622-8458
15 TELEFAX: 216-241-0816
16 INFORMATION FOR SEQ ID NO: 9:
17 SEQUENCE CHARACTERISTICS:
18 LENGTH: 443 amino acids
19 TYPE: amino acid
20 STRANDEDNESS: single
21 TOPOLOGY: linear
22 MOLECULE TYPE: peptide
23
24 US-08-653-650A-9

```

Query Match	7.2%	Score 157;	DB 2;	Length 443;
Best Local Similarity	23.6%;	Pred. No. 4.6e-08;		
Matches 95;	Conservative 50;	Mismatches 210;	Indels 48;	Gaps 13;

QY 39 QEANAALQESVPLDVTMTCNGGHWITRCQGLIEAEVDYRHFSS--CFFIPPEAGEAY 96  
 Db 52 REAGVY- EVAP-----AGPAMVITTDALAEVLADREYKVDPLDAPLAMRGVDDGL 104  
 QY 97 D-----FIPSTMDPPEOROFALANQVVMGVNDELNRIOELCSTIESL-----R 143  
 Db 105 DIPVELRPFLLIADVGDGHDNRRLRIRHAPAFNPRRLAERTDRIAALDRLITELADSSDR 164  
 QY 144 PQCGNFTEDYAEFPPIRIFMLLGLP-----EEDIPHLKYL-----TDQMTRPD 188  
 Db 165 SGEPAELLIGGAHYHFPPLVLTICELGVPVTPDPMAREAVGVKLALGLGSPAGSGDGTDP 224  
 QY 189 GSMFPAEAKELALYUFLPIICROKQPGTDAISIVANGQVNGRITSDPAKRMQGLLYG 248  
 Db 225 GDVDPDTSALESCL--LLEAVHAAKKDKDTRITMTVLYERAQAEFGVSQDOLVMTTGLIPA 262  
 QY 249 GLDTPVNFLLSTSMELAKSPENHOELIERPELLIPACEELLRRFSLY-ADGRILTSDEY 306  
 Db 283 GHDTTGSPLGF---LLEAVLGRILAADADGALISRFVEALRHHNRPPTYLLMRAAEVY 339  
 QY 307 FHGVQLKKGDDIILPQMLSGIDERRKMACPMHVDFSKQVSHTTTFGHGSHLCLQGNLARE 366  
 Db 340 IRGVALLPGAGVAVLDIEGTNDGGRHHDAPAHAFHDPDRFSRRRLTFGDGDPHYCIGQALOLE 399  
 QY 367 IIVTLKEMVLRIPDESTA-PCAQDQ--HKSSIVGVALPLVW 406  
 Db 400 SRTMIGVLRSEFPQARLAVPELRLMCRKGAQTRLDLP--VW 441

RESULT 13  
US-08-396-218-2  
: Sequence 2, Application US/08396218  
: Patent No. 5695966  
: GENERAL INFORMATION:  
: APPLICANT: INVENTI, Augusto  
: APPLICANT: BREME, Umberto  
: APPLICANT: COLOMBO, Anna L  
: APPLICANT: HUTCHINSON, Charles R  
: APPLICANT: OTTEN, Sharee  
: APPLICANT: SCOTTI, Claudio  
: TITLE OF INVENTION: PROCESS FOR PREPARING DOXORUBICIN  
: NUMBER OF SEQUENCES: 4  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: NIKAIIDO, MARNELESTEIN, MURRAY & ORAM

```

1  STREET:  655 Fifteenth Street, N. W., Suite 330 - G
2  STREET:  Street Lobby
3  CITY:    Washington
4  STATE:   DC
5  COUNTRY: USA
6  ZIP:     20005-5701
7
8  COMPUTER READABLE FORM:
9  MEDIUM TYPE: Floppy disk
10 COMPUTER: IBM PC compatible
11 OPERATING SYSTEM: PC-DOS/MS-DOS
12 SOFTWARE: Patent Release #1.0, Version #1.30
13 CURRENT APPLICATION DATA:
14 APPLICATION NUMBER: US/08/396,218
15 FILING DATE:
16
17 CLASSIFICATION: 536
18 ATTORNEY/AGENT INFORMATION:
19 NAME: KIRTS, Monica C
20 REGISTRATION NUMBER: 36,105
21 REFERENCE/DOCKET NUMBER: P1615-5002
22 TELECOMMUNICATION INFORMATION:
23 TELEPHONE: 202/638-5000
24 TELEFAX: 202/638-4810
25
26 INFORMATION FOR SEQ ID NO: 2:
27
28 SEQUENCE CHARACTERISTICS:
29 LENGTH: 422 amino acids
30 TYPE: amino acid
31 TOPOLOGY: linear
32
33 MOLECULE TYPE: protein
34
35 US-08-396-218-2

```

Query Match	7.1%;	Score 155;	DB 1;	length 422;
Best Local Similarity	23.8%;	Pred. No. 6.9e-08;		
Matches 97;	Conservative 51;	Mismatches 202;	Indels 58;	Gaps 15

QY	39	QENANVLOESVVPDLVMTKRGNGMHMIAITRQGLIEAEVEDYHNFSECCPTIPREGEAYDF	98
Db	31	REACHV- EVAP-----AGPAMVITTDALAREVLADPRFVAD-----PDLPAPAMRG	78
QY	99	IPISMD--PREOROFALA-----NQVGMENV-----KLENNIOELACSLIESL-	142
Db	79	VDDGDIDIPVLRPFTLLAGEAHRRLRIIHAFNPNRRLAEKTDIRIILAGSLIYELA	139
QY	143	-----RPOGOCNFTEDYAEPPFIRIFMLLAGLP-----EEDIPHLKYL-----TDQ	183
Db	139	DASGSGKPAELIGGFAYHNPFLVLCELLAGVPTDPMAREAVSVLKALGIGQSGGD	198
QY	184	MTRPGSMTFEAKREALYDYLPIILEORQKRGTDALISIVANGOVNRPITSDKAKMCG	243
Db	199	GTPDRGVPDTSALESL--LEAENASARNDPTPTWTRYLERAADEFGSVDDOLUYMT	256
QY	244	LLVGLGDLTVNPLFSFMEFLAKSPENHOELIERPELLPAACELLRRFSV--ADGRIL	304
Db	257	GLIFAGHNTTSSFLG---LAEVLAGLADADBDVAVSRYVEALRKHPPVPTTLMRFA	313
QY	302	TSDYEFHGVLKCKDDQILLPMLSGELDERKNACPMHNVDFSOKVSHTFCHGSHILCGH	361
Db	314	ATEWTGIGVRLPRGARVYLVDIEGTNTDGRHNDHAHFHPDRSPWRRLTFGDGPHYICGQ	373
QY	362	LAREIIVTLKEMLTRIPDFSIA-PCAQIQ--KHSGLVSGQALPLVW	406
Db	374	LAQESRTMIGVLSRPEEARLAVUYELLMRCRGAQOTARLTLEP--WV	420

RESULT 14  
US-08-760-116-2  
Sequence 2, Application US/08760116  
Patent No. 5766190  
GENERAL INFORMATION:  
APPLICANT: INVENTI, Augusto  
APPLICANT: BREME, Umberto  
APPLICANT: COLOMBO, Anna L  
APPLICANT: HUTCHINSON, Charles R

```

: APPLICANT: OTTEN, Sharee
: APPLICANT: SCOTTI, Claudio
: TITLE OF INVENTION: PROCESS FOR PREPARING DOXORUBICIN
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSER: NIKAIKO, MARCELSTEIN, MURRAY & ORAM LLP
: STREET: 655 Fifteenth Street, N. W., Suite 330 - G
: CITY: Washington
: STATE: DC
: COUNTRY: USA
: ZIP: 20005-5701
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/760,116
: FILING DATE: 3-DEC-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/396,218
: FILING DATE: 27-FEB-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: KITTIS, Monica C
: REGISTRATION NUMBER: 36,105
: REFERENCE/DOCKET NUMBER: P1615-6007
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202/638-5000
: TELEFAX: 202/638-4810
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 422 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-760-116-2
```

```

Query Match          7.1%; Score 155; DB 1; Length 422;
Best Local Similarity 23.8%; Pred. No. 6.9e-08;
Matches 97; Conservative 51; Mismatches 202; Indels 58; Gaps 15;

QY 39 OEAMAVLQESNVPLDVTWTRCGMHIAIRGQILREAVEDYHFSSECPFIPREAGEAYDF 98
DB 31 REAGPVV-EVNAP-----AGPRAMVITDOLAREVLADPRFVAD-----PDLPAMARG 78

QY 99 IPTSMD--PPROKOFRLA-----NOVGMFVVD-----KLENRIOELACSLIESL- 142
DB 79 VDDGIDIPVPELRFETLIVADGEAHRRLRIHAPFNPRRLAERTDRIALAGRLTLELA 138

QY 143 ---RPGQCFNTEDEVAPFPPIRIFEMLLAGP-----EEDIPHLVYL-----TDQ 183
DB 139 DASRSCKPAELIGFAVHPFLVLYICELGVPTDPMAREAVSLALGIGRQSGGD 198

QY 184 MTRPDGSMTEFAEAKALYDYLPIIEQROKPGTDAISIVANGVNGRPTISDEAKRMCG 243
DB 199 GTDPAGGVPRDSALESL--LLEAVHSARNDPTMTFVLYERAAGFSGVSDDDLVYMIT 256

QY 244 LLLVGGLDVTYVNLFSFSEKFLAKSPENHROELIERPELIPACCELLRRFSLV--ADGRIL 301
DB 257 GLTFAGHDTGSGFLGF--LLAEVLAGRLAADADEDAVSREVEELRYHPPVPTLWRA 313

QY 302 TSDVEFGVYOLKKGDDLLPQMLSGLDERKACPMHVDFSROKYSHTTFGSHLCLGQH 361
DB 314 ATEVTTIGVRLPRGAPVAVLDIEGNTDGRHNDAPHAPDRPSRRRLTFGGDPHYCIGEQ 373

QY 362 LARREIIVTLKEMLTIRIPDSIA-PGAQIQ--HKSGIVSGVQALPLVW 406
DB 374 LAQESHTMIGVLSRFPPEARLAVPYDELRLMCRKGAGTARLTLEP-VW 420
```

```

RESULT 15
US-08-194-981E-5
: Sequence 5, Application US/08194981E
: Patent No. 5886157
: GENERAL INFORMATION:
: APPLICANT: GUENGERICH, F. Peter
: APPLICANT: GUO, Zuyu
: APPLICANT: SANDHU, Punam
: APPLICANT: GILLAM, Elizabeth M. J.
: TITLE OF INVENTION: EXPRESSION AND PURIFICATION OF
: TITLE OF INVENTION: HUMAN
: NUMBER OF SEQUENCES: 68
: CORRESPONDENCE ADDRESS:
: ADDRESSER: NEEDLE & ROSENBERG, P.C.
: STREET: Suite 1200, 127 Peachtree Street, NE
: CITY: Atlanta
: STATE: Georgia
: COUNTRY: USA
: ZIP: 30303-1811
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/194,981E
: FILING DATE: February 10, 1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Elizabeth Selby
: REGISTRATION NUMBER: 38,298
: REFERENCE/DOCKET NUMBER: 22000.0022
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (404) 688-9880
: TELEFAX: (404) 688-9880
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 512 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: HYPOTHETICAL: NO
: FRAGMENT TYPE: N-terminal
US-08-194-981E-5
```

```

Query Match          6.2%; Score 135; DB 2; Length 512;
Best Local Similarity 22.4%; Pred. No. 1.2e-05;
Matches 95; Conservative 67; Mismatches 158; Indels 104; Gaps 21;

QY 27 DMVPSNLSGVQAMAVLQESNVPLDVTWTRCGMHIAIRGQILREAVEDYHFRF----- 81
DB 107 DLYFTLLISNGQSMFSF-----PD-----SGPVAAAR-----RLAONGLSFSIASD 149

QY 82 ---SSCEPFIPEAGEAYDFIPTSM---PPROKOPR-----ALANOVGMFVVDKLENR 130
DB 150 PASSTSCYLEHYSKEAEVLISTIQLMAGRGHFNPRVYVASTVNICACGRARDHN 209

QY 131 IOELACSLIESLRPGQCFNTEDEVAPFPPIRIFEMLLAGLPEEDIPHLKYLTDOMTRPDGS 190
DB 210 HQEL-LSLV-----NINNNNGE-----VSGSNPAEFIPILRYL-----PNPS 246

QY 191 M-TFAEAKALYDYLPIIEQROK-----KEGTDAISIVANGQ-----VNGRPTISDE 237
DB 247 LMAEKDLINEFYSFMQMKVAKHKKTFEKGHIRITTD--SLIEHQEKQOLDENANVOLSD 304

QY 238 AKRMCGLLVG-GIDYVNLFSFSEKFLAKSPENHROELIE-----RPEL----- 280
DB 305 KIINIYVLDLGAGFDYTTAISMSIMYLVANPRQKRIQDELDTVIGRSRRPLSDRSHL 364

QY 281 -IPACEELLRRFSLV--ADGRILTSDEYFGVYOLKKGDDILLPQMLSGLDERKACPM- 335
```



Db 365 PYMEAFLEFFRHSSFVFTIPHSSTRDTSIKGFYIPKGRCVFNQWQINHDOKLWNP 424  
QY 336 -----MHVDSROKV---SHTFGSHLCLGOHLARREIYTLKEMLTRIPDFSIAP 385  
Db 425 EELPERFLTPDGAIDKVLSEKVIIFGMRKRCIGETVARWEVFLLAILLQRY-EFSVPL 483  
QY 386 GAQI 389  
Db 484 GVKV 487

Search completed: October 4, 2000, 12:38:07  
Job time: 21424 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: October 4, 2000, 12:36:41 ; Search time 73.09 Seconds  
(without alignments)  
350.531 Million cell updates/sec

Title: US-09-246-451-12

Perfect score: 2179

Sequence: 1 TTEIQSNANLALPPHPVE.....IVSGVQALPLVMDPATKAV 414

## Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 178050 seqs, 61884766 residues

Total number of hits satisfying chosen parameters: 178050

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

## Database :

1: PIR-64:\*  
2: pir1:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2169	99.5	415	1	campfor 5-monoxyg
2	378	17.3	410	2	cytochrome P450meg
3	374	17.2	393	1	cytochrome P450 Rv
4	372	17.1	396	1	cytochrome P450 yj
5	370	17.0	397	1	cytochrome P450 my
6	366	16.8	398	1	cytochrome P450 Rv
7	352.5	16.2	404	2	cytochrome P450 -
8	346	15.9	403	2	cytochrome P450 10
9	342.5	15.7	428	1	cytochrome P450 Rv
10	341.5	15.7	410	1	cytochrome P450 10
11	339.5	15.6	405	1	cytochrome P450 CV
12	339	15.6	410	1	cytochrome P450 cy
13	336	15.4	406	2	cytochrome P450 10
14	332	15.2	406	1	cytochrome P450 10
15	327.5	15.0	376	1	polyketide synthase
16	327	15.0	411	1	polyketide synthase
17	324	14.9	428	1	cytochrome P450ter
18	322.5	14.8	395	1	cytochrome P450 bl
19	319.5	14.7	410	2	cytochrome P450 sc
20	318.5	14.6	399	1	cytochrome P450 -
21	317.5	14.6	405	1	cytochrome P450 Rv
22	315	14.5	412	2	cytochrome P450 (s
23	311	14.3	310	2	cytochrome P450 (i
24	310	14.2	438	1	cytochrome P450 Rv
25	309	14.2	402	2	cytochrome P450 Rv
26	309	14.2	404	1	cytochrome P450 Rv
27	309	14.2	417	1	cytochrome P450 Rv
28	302.5	13.9	408	2	cytochrome P450 no
29	302.5	13.9	408	2	cytochrome P450 no

30	293.5	13.5	406	1	cytochrome P450 er
31	293	13.4	381	1	cytochrome P450 cy
32	291.5	13.4	433	1	cytochrome P450 Rv
33	290.5	13.3	337	2	cytochrome P450 hy
34	273.5	12.6	310	2	probable hydroxyla
35	271	12.4	414	1	cytochrome P450 Rv
36	268.5	12.3	411	2	probable cytochrom
37	267.5	12.3	386	2	cytochrome P450
38	259	11.9	401	1	cytochrome P450 BJ
39	259	11.9	412	1	cytochrome P450 BJ
40	255	11.7	411	1	cytochrome P450 -
41	254	11.7	406	3	cytochrome P450 hy
42	240	11.0	396	1	cytochrome P450 Rv
43	233	10.7	433	2	cytochrome P450 ho
44	231	10.6	398	1	vitamin D-3 25-hyd
45	229	10.5	402	1	cytochrome P450 Rv

## ALIGNMENTS

```
RESULT 1
O4PSCP
campfor 5-monoxygenase (EC 1.14.15.1) cytochrome P450 101 - Pseudomonas putida plasm
N:Alternate names: cytochrome P450-CAM
C:Species: Pseudomonas putida
C>Date: 30-Apr-1982 #sequence-revision 31-Dec-1993 #text-change 03-Mar-2000
C:Accession: A25660; S34614; C60886; A00194
J:Unger, B.P.; Gunsalus, I.C.; Sliagar, S.G.
J: Biol. Chem. 261, 1158-1163, 1986
A:Title: Nucleotide sequence of the Pseudomonas putida cytochrome P-450-cam gene and
A:Reference number: A94678; MUID:86111751
A:Accession: A25660
A:Molecule type: DNA
A:Residues: 1-415 <UNG>
A:Cross-references: GB:M12546; MID:g151114; PIDN:AAA25760.1; PID:g151115
R:Atamaki, H.; Koga, H.; Sagar, Y.; Hosoi, M.; Horinuchi, T.
Biochim. Biophys. Acta 1174, 91-94, 1993
A:Title: Complete nucleotide sequence of the 5'-exo-hydroxycampfor dehydrogenase gene
A:Reference number: S34613; MUID:93326643
A:Accession: S34614
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-42 <ARA>
A:Experimental source: PG3; ATCC 17453; CAM plasmid
R:Romeo, C.; Moriwaki, N.; Yasunobu, K.T.; Gunsalus, I.C.; Koga, H.
J. Protein Chem. 6, 253-261, 1987
A:Title: Identification of the coding region for the putidaredoxin reductase gene fro
A:Reference number: A60886
A:Accession: C60886
A:Molecule type: DNA
A:Residues: 408-415 <ROM>
R:Haniy, M.; Armes, L.G.; Yasunobu, K.T.; Shastri, B.A.; Gunsalus, I.C.
J. Biol. Chem. 257, 12664-12671, 1982
A:Title: Amino acid sequence of the Pseudomonas putida cytochrome P-450. II. Cyanogen
A:Reference number: A00194; MUID:83030788
A:Accession: A00194
A:Molecule type: protein
A:Residues: 2-55, 58-276, 'Q', 278-361, 'S', 363-407, 'N', 409-415 <HAN>
C:Genetics:
A:Gene: camC; CYP101
A:Genome: plasmid
C:Function:
A:Description: catalyzes hydroxylation of campfor to yield 5-exo-hydroxycampfor; elec
C:Superfamily: Pseudomonas plasmid campfor 5-monoxygenase; cytochrome P450 homology
C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygena
F:246-380/Domain: cytochrome P450 homology <CYP>
F:358/Binding site: heme iron (Cys) (axial ligand) #status experimental

Query Match 99.5%; Score 2169; DB 1; Length 415;
Best Local Similarity 99.5%; Pred. No. 2,9e+158;
Matches 412; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```

QY 1 TTTTQSNANLAPLPHPYHEHLVDFDMYNPSNL.SAGVOEAMAVLQESNVDLWTRCNG 60
    |||||
Db 2 TTTTQSNANLAPLPHPYHEHLVDFDMYNPSNL.SAGVOEAMAVLQESNVDLWTRCNG 61
QY 61 GHWIATGQGLREAYEDYRHSSCEPFIIPRAGAYDFIPISMPDPDROQRALANOVYG 120
    |||||
Db 62 GHWIATGQGLREAYEDYRHSSCEPFIIPRAGAYDFIPISMPDPDROQRALANOVYG 121
QY 121 MPVVDKLENRIQELACSLIESLRPOGOCNFEEDYAEPPRIIFMLAGLPEDIPHLKYL 180
    |||||
Db 122 MPVVDKLENRIQELACSLIESLRPOGOCNFEEDYAEPPRIIFMLAGLPEDIPHLKYL 181
QY 181 TDQMTPRDGSMTFAEAKALDYLIPIIEORRQKPGTDAISIVANGOVNGRPITSDEAKR 240
    |||||
Db 182 TDQMTPRDGSMTFAEAKALDYLIPIIEORRQKPGTDAISIVANGOVNGRPITSDEAKR 241
QY 241 MCGLLVGLDVTYVNFISFMEFLAKSPENHQLIERELIPACBELLRFSLVADGRI 300
    |||||
Db 242 MCGLLVGLDVTYVNFISFMEFLAKSPENHQLIERELIPACBELLRFSLVADGRI 301
QY 301 LTSDFEFGVOLKKGDDIILPQMLSGDERKNACPMHYDFSROKVSHTTFGSHLCLGQ 360
    |||||
Db 302 LTSDFEFGVOLKKGDDIILPQMLSGDERKNACPMHYDFSROKVSHTTFGSHLCLGQ 361
QY 361 HLAAREIIVTLKEMLFRIPDFSIAPGAQIOHKSIGVSGVALPLVWDPATTAKAV 414
    |||||
Db 362 HLAAREIIVTLKEMLFRIPDFSIAPGAQIOHKSIGVSGVALPLVWDPATTAKAV 415

RESULT 2
cytochrome P450meg - Bacillus megaterium (ATCC 13368)
C:Species: Bacillus megaterium
A:Variety: ATCC 13368
C:Date: 07-Oct-1994 #sequence_revision 23-Feb-1996 #text_change 04-Mar-2000
C:Accession: S39924; S32216
R:Rauschenbach, R.; Isernhagen, M.; Noeske-Jungblut, C.; Boldol, W.; Stewert, G.
Mol. Gen. Genet. 241, 170-176, 1993
A:Title: Cloning sequencing and expression of the gene for cytochrome P450meg, the steroid
A:Reference number: S39924; MUID:94049677
A:Accession: S39924
A:Molecule type: DNA
A:Residues: 1-410 <RADU>
A:Cross-references: EMBL:221972; NID:g288298; PIDN:CAA79985.1; PID:g288300
C:Experimental source: ATCC 13368
C:Superfamily: Bacillus cytochrome P450 Cyp106; cytochrome P450 homology
C:Keywords: chromoprotein; heme; iron; metalloprotein
F:355/Binding site: heme iron (Cys) (axial ligand) #status Predicted

Query Match 17.3%; Score 378; DB 2; Length 410;
Best Local Similarity 30.4%; Pred. No. 2,1e-21;
Matches 110; Conservative 53; Mismatches 179; Indels 20; Gaps 7;

QY 63 WIATGQGLREAYEDYRHSSCEPFIIPRAGAYDFIPISMPDPDROQRALANOVYG 115
    |||||
Db 44 WNVFKYEDKRVSLDYKHFSVRRKRTTISVGTDESEGSVEKIDITSDPPDRHRRSL 103
QY 116 NQVVGMMVVDKLENRIQELACSLIESLRPOGOCNFEEDYAEPPRIIFMLAGLPEDIP 175
    |||||
Db 104 AAATPSPSLNMEPRIOELDELIGMDGTEIDIVASLSPRIYMAIDMGVPSKDR 163
QY 176 HLKYLFDQMTRPDGSMTFAE-----AKBALYDYLIPITIEORRQKPGTDAISIVANGOV 228
    |||||
Db 164 LFKKWDITLFLPDREKQOEVDKIKOYAAKEYOYLKPIYVQKLANADDIISLAKSEV 223
QY 229 NGRPITSDEAKRMCGLLVGLDVTYVNFISFMEFLAKSPENHQLIERELIPACBEE 287
    |||||
Db 224 DGMFTFDEVDVARTMLITGAVETTSHLANSFSLYDKEVYQELHENLDLVQAVEE 283
QY 288 LLR-RFSLVADGRIITSDFEFGVOLKKGDDIILPQMLSGDERKNACPMHYDFSROKVSHTTFGSHLCLGQ 345
    |||||
```

```

Db 284 MLREFMLIKIDRTVKEDNDLGLVELKEDSVVVMSSAANNDMEFDPFTLNIHRPNK 343
QY 346 SHTFEGHSHLCLGQHLARREIIVTLKEMLFRIPDFSIAPGAQIOHKSIGVSGVAL 402
    |||||
Db 344 KHLTFGNGPHRCGLAPLARLEAKIALTAFLKFKHIEAVPSFOLENLTDTSATGOTLSL 403
QY 403 PL 404
    ||
Db 404 PL 405

RESULT 3
cytochrome P450 rv1785c - Mycobacterium tuberculosis (strain H37Rv)
N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Mycobacterium tuberculosis
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: C70929
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Field, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sultson, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A:Reference number: A70500; MUID:98295987
A:Accession: C70929
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-393 <COL>
A:Cross-references: GB:AL022021; GB:AL123456; NID:g3250699; PIDN:CAA17707.1; PID:e125
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: rv1785c
C:Superfamily: Bacillus cytochrome P450 Cyp106; cytochrome P450 homology
C:Keywords: oxidoreductase

Query Match 17.2%; Score 374; DB 1; Length 393;
Best Local Similarity 28.7%; Pred. No. 4e-21;
Matches 110; Conservative 69; Mismatches 184; Indels 20; Gaps 10;

QY 31 PSNLSAGVOEAMAVLQESNVDLWTRCNGHWIATGQGLREAYEDYRHSSCEPFIIPR 90
    |||||
Db 23 PMAYDRGV--GWKTLRDAG--PVVF--MNGWYTL-TRREVDLAALRNPXYFSSRKALQP- 74
QY 91 EAGCAYFIPTSMPPDROQRALANOVGMPPVDKLENRIQELACSLIESLRPOGOCNF 150
    |||||
Db 75 -PGNPLVVPLADPPRPHTRIRRILOPYFSRPAALSKALPSLRRTYVAMIDAIAGRGCEA 133
QY 151 TEDIAEPPRIIFMLAGLPEDIPHLKYLTDQM---TRPDGSMFAEAKALDYDLIP 206
    |||||
Db 134 MADLANLFPPOLFLVILGLLEDRDLIGKDAVIAMSDRPHTEADVAARLELLETLTA 193
QY 207 IIEORRQKPGTDAISIVANGOVNGRPITSDEAKRMCGLLVGLDVTYVNFISFMEFLAK 266
    |||||
Db 194 MVAERRRNPGPDVLSQV---QIGEDPLSEIEVLGSHLILAGIDVTAAVGSLSLELAR 250
QY 267 SPENHQLIERELIPACBELLR-RFSLVADGRIITSDFEFGVOLKKGDDIILPQMLS 325
    |||||
Db 251 RPQRLAARLNPKQIRVFETIEIVALEPSAVAPRVTEPTVGGMTLPASSPVRLCAAV 310
QY 326 GLDERKNACPMHYDFSROKVSHTTFGSHLCLGQHLARREIIVTLKEMLFRIPDFSIAP 385
    |||||
Db 311 NRQSDAMSTDELVMQDKVHRHMGFGGPRKRCGSHLARLELTLVGEWLNQIPLDFELAP 370
QY 386 --GAQIOHKSIGVSGVALPLVW 406
    |||||
Db 371 DYAPEIRFPSKSA-LKNLPLRW 392

RESULT 4
cytochrome P450 yj1b - Bacillus subtilis
```



Matches	108;	Conservative	69;	Mismatches	168;	Indels	34;	Gaps	8;
QY	55	WTRCN-----	GGHMIAITRCQLIREAEDYRHSS-----	ECFPIREAGEAYDIPT	101				
Db	26	WMRANQPVFPRBNGLAAASTYQAVIDAEROPELFFSNAGGIRPDQPALP-----	MMI	76					
QY	102	SMDDPEQGFALANQVYGMVYDKLENRIQELACLSIESLRPGQGCNFTEDYAEFFIR	161						
Db	77	DMDDDAHLIRKRLVYAGTTRKRKVKDKDEASTALCPTLLIDACERECDFVBDLAPLPMA	136						
QY	162	IFMLLAGL-PEEDIPFHLK-----	YLTQDMTRPDGSMTFPAEAKALDYILPIIQRQ	213					
Db	137	YIGDMLGVRPEGRQMFELKMSDDLYTFELSSHVSQEDFQITM-DAPFAYNDPFRATIAARA	195						
QY	214	KPQDAISIVANGOVNRPITSDEAKRMCGILLVGCGLDTVNVNLSFESMEFLAKSPENHOE	273						
Db	196	PTDIDLVSILVSEVDGSRSLSDDELVMETLLILLIGGDETFRHTLSGTEQQLIRNDDWL	255						
QY	274	LIERELLIPACCELLRRFSLYAD-GRITSDYERHGVQALKKQOILLPQMLSGLDEKN	332						
Db	256	LORDSLIPGALIEELMLRTAPKKNCRVLTADTEHGTALAGEKMMILFESANDEAVF	315						
QY	333	ACPMHVDSESRQVSHHTTGHGSHLCLGQHLARRELIIVLKLEMLTIRPPFSAIPAQOI--Q	390						
Db	316	CEPEKFDVQQRNNSHILARGFTHFCLGQNLARLDELSTMTENYLRRLPLRLVADVDSVLP	375						
QY	391	HKSGTIVSGVALPLVMDPA	409						
Db	376	RPANFVSGLESMPVVFPTPS	394						

```

RESULT 7
t30231
cytochrome P450 - Streptomyces hygroscopicus
C:Species: Streptomyces hygroscopicus
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 31-Jan-2000
C:Accession: t30231
R:Apiciclov, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Staun
Gene 169, 9-16, 1996
A:Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hyg
A:Reference number: 220782
A:Accession: t30231
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-404 <A>A>
A:Cross-references: EMBL:X86780; NID:g987088; PID:g987105; PIDN:CAA60465.1
C:Genetics:
A:Gene: rapN
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology

Query Match 16.2%; Score 352.5; DB 2; Length 404;
Best Local Similarity 28.9%; Pred. No. 1.8e-19;
Matches 103; Conservative 63; Mismatches 176; Indels 15; Gaps 9;

QY 63 WATGQGLIREVYEDYRHFSSCFP---IPREAGATYF-IPTSMDPPEQGFRLANQV 118
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 50 WLAASMEVAVYFVDPF-FRSNATLGKDPVRLPAIQQPVIMLMDPEHRLRLRVAKRA 108
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 119 VGMPPVDKLENRIQELACSLIESLRQG-QCNFTEVAYEPPRIIFMLAGLPEDIPHL 177
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 109 LTSRMELALRRTOGVADDLIDKMLAKAPADLMEDFALPLIIMICGLGVPIEDQTFK 168
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 178 KYLDQMTPRPGSMT---FAEAKELDYLLPIIEQRQKGTDAISIVANGVNGRPIT 234
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 169 RTWSQDML-SNGAYSQEVYMAAGOSLYLLELILAEIRKKQOTNDLGLSVARAKDRDRLS 227
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 235 SDEAKRMGCLLVGGDLTVNVEFLSFMELAKSPREHROELERPELIPAAACEELLRRSL 294
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 228 ETELVGFAVTLILINGVETTTAANGSVYTLTLTHPEKLAELKRLDLSIPKAVDELIRIPI 287
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 295 VADG---RLTSDVEEFHGVLAKGQDILLPQMLSGLDERKNACPMHWDFSRQKVSHTEFG 351
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Db 268 AKQASWPMVAIVEDSELSTLYKAGEAVALIQHSANTDPKYVDHPEIDFHRTSNPHMSLG 347

QY 352 HGSHLCTLOOHARREIIVTLKEWLIRIPDPSFA-PGAOIQHKSQ-IVSGVQALPLW 406  
||| | | | : | : |||| | : | : |||| |  
Db 348 HGHHHCMAQLVRVMQTALGSLISRIPLRFLVAVPRIKRLRGRLVLSLEALLPLTW 404

RESULT 8

B35401  
cytochrome P450 105B1 - Streptomyces griseolus  
N:Contains: oxidoreductase (EC 1.-.-.-)  
C:Species: Streptomyces griseolus  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 04-Mar-2000  
C:Accession: B35401  
R:Omer, C.A.; Lénstra, R.; Little, P.J.; Dean, C.; Tepperman, J.M.; Leto, K.J.; Rommess  
J. Bacteriol. 172, 3335-3345, 1990  
A:Title: Genes for two herbicide-inducible cytochromes P-450 from Streptomyces griseo  
A:Reference number: A35401; MUID:90264332  
A:Accession: B35401  
A:Molecule type: DNA  
A:Residues: 1-403 <OME>  
A:Cross-references: GB:M32239; NID:g153480; PIDN:AAA26825.1; PID:g153481; GB:M36481  
C:Genetics:  
A:Gene: CYP105B1  
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology  
C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygenase  
f:236-374/domain: cytochrome P450 homology <P45>  
f:352/Binding site: heme iron (Cys) (axial ligand) #status predicted

```
Query Match      15.98; Score 346; DB 2; Length 403;  
Best Local Similarity    29.28; Pred.No. 5.7e-19;  
Matches 106; Conservative   63; Mismatches 164; Indels   30; Gaps   13;
```

Dy

QY 63 WATFGQLIREAYEDYRHFSESC-----PFIPEGAVDYDFIPT--SMDPEOROFALIA 115  
| : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 52 WLVTFHODVRAVLGDPR-FSDAHRHTGPFLLTAGSGREILGNPTFLRMDDPEHARLRML 110  
  
QY 116 NOVGMFPVVDKLENRIQELACSLIESLNP-OQCNCFTDVAEPPPIRFEMLAGIPEDI 174  
| : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 111 TADFIVKKVEAMREPVOGLADDLVDRMTTGTSSADVLETFAPLPISLYICILLVVVEDYA 170  
  
QY 175 PHL-----KYLDQMTRPGSMTEFAEKALKVDYLPIITEGRORKPDGIS-IYANGOVN 229  
| : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 171 AFPOERSNVLLTLNSTPE---EVRAAOBELLEYLARLKARTKRERDDAILSRIVARGELD 227  
  
QY 230 GRPITSDEAKMCGCLLVGLGDTPVNPLSFMSMEPLANSPEHHOLIERPELLPAACEELL 289  
| : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 228 DTQIANT----MGRLILAAGHETTANMATLTLYLTRNDPOLARAEPALVKAVEELL 282  
  
QY 290 RRESLIYVAGC-RILTSDVEPHGVOLKKGDOLLPMISMIGDERKNACC-MHVDESROKV 345  
| : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 283 RYLITVNHGVPRiatedvLGRTIAGEGYLC--MISSARRDAEVPFGDDLVARPAR 340  
  
QY 346 SHTEGHGSHELCTGOHLAREIIYVLKEMLTRIIDFSIA-PGAOIOHKSCI-VSVGVALP 403  
| : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 341 RHVAFGEFVHCLOGCPALARVELQAIEFLLRRLEDRLAIRAHBEIELFRGGMAITYGHSHP 400  
  
QY 404 LVW 406  
| : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 401 IAW 403

RESULT 9  
F70729  
cytochrome P450 RV2266 - Mycobacterium tuberculosis (strain H37RV)  
N:Contains: oxidoreductase (EC 1.-.-.)  
C:Species: Mycobacterium tuberculosis  
C>Date: C10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
CAccession: F70729  
RCole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,  
J.; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
Bajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, S.F.; Skelton, S.; Squares, S.



```
Db 54 QALLD-----PRLAKDGRTOQIIEKRIADAERRPGFSPDLGPHMLTDPDHTRLR 105
113 ALNOVYGVMPVVDKLENRIOELACSLIESLRPGOCNFTEDYAEPPRIPIFMLELAGREE 172
Db 106 KLVVKAFTARVYDGLRPIRQIOTDDLDRLAGSEVDLIDEPFPLITYISELMGEVDS 165
173 DIPHLKYLTDQMTRPDGSMTFAEAKF--ALYDYLPIIEQRORPGTDAISIANGOVNG 230
166 RRDDFSRMTNVLV--DQSGPEAQOASVAMVEXYLTLLIAKRTPEGGDILLTALLEAVED 223
231 RPTSPDARKMCGLLVGGIDTVYNFLSFMELASPEHROELIERPELIPACEELR 290
224 DRISEGELIAMVFLLVAGHETVNLIGNCVSLGNPDQALALRNDPSLLPACIEETLR 283
291 RFLVADG--RLITSDYEFHGVOLKKGDQILLPQMLSGDERKNACPMHVDSPKOVKSH 348
284 YESPVANGTFRHRAEAVREDVDYIPGELVYVVALGAAANRGEFEDPRDRTIRETTGHV 343
Db 349 TFGHSHLCLGOHLARREIIVTLKEMLTRIPDSIA--PGAQIQHKSIGVGVQALPL 404
344 AFCHGHFPCGALALAEAGIANGRLERPRDLMAASPDLLMRFSLVLRGIEKLPLV 401
```

## RESULT 12

```
cytochrome P450 CYP4 - Bacillus subtilis
N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Bacillus subtilis
C.Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Feb-2000
C.Accession: E69611; 144774
R:Kumai, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choc A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Filtz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier, J.; Hamwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Koester, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, K.; Liu, H.; Masuda, S.; Maueel, Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetalle, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, A.; Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron, A.; Whiter, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Whiter, M.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasunori, K.; Yata, K.; Yoshida, K. A:Authors: Yoshikawa, H.F.; Zumschein, E.; Yoshikawa, H.; Danchin, A. A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis. A:Reference number: A69580; MUID:98044033
A:Accession: E69611
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-410 <KUN>
A:Cross-references: GB:299117; GB:AL009126; NID:g2634966; PIDN:CAB14615.1; PID:el183903; A:Experimental source: strain 168
R:Belitsky, B.R.; Gustafsson, M.C.U.; Sonenshein, A.L.; von Wachenfeldt, C. J. Bacteriol. 179, 5448-5457, 1997
A:Title: An lrp-like gene of Bacillus subtilis involved in branched-chain amino acid tra A:Reference number: Z22837
A:Accession: T44774
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-410 <BEL>
A:Cross-references: EMBL:Y11043; PIDN:CAV1937.1
A:Experimental source: strain 1A1
C:Genetics:
A:Gene: cypA
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C:Keywords: oxidoreductase
E:245-381/Domain: cytochrome P450 homology <CYP>
```

```
Query Match 15.6%; Score 339; DB 1; Length 410;
Best Local Similarity 27.5%; Pred. No. 2e-18;
Matches 87; Conservative 69; Mismatches 138; Indels 22; Gaps 5;
```

```
QY 104 DPEOROFRALANOVGMVVDKLENRIOELACSLIESLRPGOCNFTEDYAEPPRIPIF 163
113 ALNOVYGVMPVVDKLENRIOELACSLIESLRPGOCNFTEDYAEPPRIPIFMLELAGREE 172
Db 100 DPEOROFRALANOVGMVVDKLENRIOELACSLIESLRPGOCNFTEDYAEPPRIPIF 159
164 MLAGPEEDIPHLKYLTDQMTRPDGSMTFAEAKF--ALYDYLPIIEQRORPGTDAISIANGOVNG 230
160 SEMIGIPLEDROKFRVWSQA-----IDFSDAERLQENDHLLGEFEVLESLVKKRR 213
214 KPGTDVSIYANGOVGRPTSPDARKMCGLLVGGIDTVYNFLSFMELASPEHROELIERPELIPACEELR 290
214 EPAGDLISALIQAESGTOISTEELYSMTLLVAGHETVNLITMGTVALMCHDQLEK 273
QY 274 LIERPELIPACEELIRFSLV--ADGRILTSYEFHGVOLKKGDQILLPQMLSGDERKNACPMHVDSPKOVKSH 348
274 LRQPDLMNSAIEALRPHSPVELTTLIRWAEFFILAGOEIKKKDVIIISLASANDEK 333
Db 332 NACPMHVDSPKOVSHTTFGHSHLCLGOHLARREIIVTLKEMLTRIPDSIA--PGAQIQHKSIGVGVQALPL 404
334 FPNADIFDIERKNRRIRIAFGHGNFCLGAOLARLEAKIAISTLLRRCPTQL-KGEKKQM 392
QY 392 K---SGIVSGVQALPL 404
393 KWKGNFLMRALIELPLI 408
```

## RESULT 13

```
cytochrome P450 105A1 - Streptomyces griseolus
N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Streptomyces griseolus
C.Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 04-Mar-2000
C.Accession: A35401
R:Omey, C.A.; Lenstra, R.; Little, P.J.; Dean, C.; Tepperman, J.M.; Leto, K.J.; Romess J. Bacteriol. 172, 3335-3345, 1990
A:Title: Genes for two herbicide-inducible cytochromes P-450 from Streptomyces griseo A:Reference number: A35401; MUID:90264332
A:Molecule type: DNA
A:Residues: 1-406 <OMF>
A:Cross-references: GB:M32238; NID:9153477; PIDN:AAA26823.1; PID:9153478; GB:M36480
C:Genetics:
A:Gene: CYP105A1
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygenase E:241-377/Domain: cytochrome P450 homology <P45>
F:355/Binding site: heme iron (Cys) (axial ligand) #status predicted
```

```
Query Match 15.4%; Score 336; DB 2; Length 406;
Best Local Similarity 26.6%; Pred. No. 3.3e-18;
Matches 105; Conservative 74; Mismatches 192; Indels 24; Gaps 11;
QY 31 PSLMSAGVQ--EAMAVLQESNPVDLWTRKNGH-WIARIGQLIRAYEDYR----- 79
113 ALNOVYGVMPVVDKLENRIOELACSLIESLRPGOCNFTEDYAEPPRIPIFMLELAGREE 172
Db 17 PNRSRCPYQLPDQYQDLRTPGRLHVTVLYDGRQAVVWVHNEAARKILDPRLSNRTDD 76
80 HPSSECPPIP--REAGEAVDPIPTSMDDPEQGRFRALANOVGMVVDKLENRIOELAC 137
77 NPFAIYSPREAVRESFOA-----IGDPPDHGTRRRMTTSEFVKKIKGRPREVEVNG 132
QY 138 LIESLRPG--QCNTEDYAEPPRIPIFMLELAGPEEDIPHLKYLTDQMTRPDGSMTFAE 196
113 ALNOVYGVMPVVDKLENRIOELACSLIESLRPGOCNFTEDYAEPPRIPIFMLELAGREE 172
Db 133 FLDEMLAAGPTADLVGQFLVPVSWICRLGLGVYADHDEFQDASRLVOSTDAQALNA 192
197 KEMLYLPIIEQRORPGTDAI--SIYANGOVGRPTSPDARKMCGLLVGGIDTVYN 255
193 RNDLADYLDGLTQFQTEGAGLVGALVADQLANGE-IDREELISTAMLLIAGHETTS 251
256 FLFSMEPLAKSPENHROELIERPELIPACEELIRFSLV--VADGRILTSYEFHGVOLK 313
252 MTSLSVITLLDHPEDYQALRADRSIVGAVEELRLAIADNAGVATADIEVSHLIR 311
```





---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 13:04:52 ; Search time 27.02 Seconds

(without alignments)  
474.815 Million cell updates/sec

Title: US-09-246-451-12

Perfect score: 2179

Sequence: 1 TTERIQSNANLAPLEPPHYPE.....IVSGVQALPLVWDPAATKAV 414

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 85661 seqs, 30989116 residues

Total number of hits satisfying chosen parameters: 85661

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_38.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2169	99.5	414	1 CPXA_PSEPU	P00183 pseudomonas
2	584.5	26.8	414	1 Y4VG_RHISN	053215 rhizobium s
3	378	17.3	410	1 CPXM_BACME	006069 bacillus me
4	374	17.2	393	1 YH85_MYCTU	053936 mycobacteri
5	372	17.1	396	1 Y71B_BACSU	034374 bacillus su
6	366	16.8	398	1 YZ1B_MYCTU	053563 mycobacteri
7	358.5	16.5	405	1 CPXM_BACSU	P27632 bacillus su
8	346	15.9	402	1 CPXE_STRGO	P18327 streptomyc
9	342.5	15.7	428	1 YM6E_MYCTU	050696 mycobacteri
10	341.5	15.7	410	1 CPXI_SACME	P14762 bacillus me
11	339.5	15.6	405	1 CPXK_SACER	P33271 saccharopol
12	339	15.6	410	1 CPXI_BACSU	008469 bacillus su
13	336	15.4	405	1 CPXE_STRGO	P18327 streptomyc
14	334	14.9	428	1 CPXL_PSESP	P33006 pseudomonas
15	322.5	14.8	395	1 BIOL_BACSU	P53554 bacillus su
16	318.5	14.6	399	1 FASI_RHOFA	P46373 rhodococcus
17	317.5	14.6	405	1 YC5E_MYCTU	011062 mycobacteri
18	315	14.5	412	1 CPXH_STRGR	P26911 streptomyc
19	310	14.2	402	1 Y180_MYCTU	008464 mycobacteri
20	309	14.2	408	1 NOR_FUSOX	P23295 fusarium ox
21	302.5	13.9	408	1 NOR2_CYLTO	012599 cylindrocac
22	301	13.8	381	1 CPXG_STRSO	P23296 streptomyc
23	294.5	13.5	403	1 NOR1_CYLTO	000616 cylindrocac
24	291.5	13.4	433	1 YZ4S_MYCTU	P71856 mycobacteri
25	288.5	13.2	404	1 CPXJ_SACER	P00441 saccharopol
26	285	12.4	400	1 CPXK_RHISN	P55544 rhizobium s
27	271	12.4	414	1 Y778_MYCTU	P77903 mycobacteri
28	270.5	12.4	397	1 CPXO_SACER	P48655 saccharopol
29	259	11.9	401	1 CPXE_BRAJA	059203 bradyrhizob
30	240	11.0	396	1 YW6E_MYCTU	059571 mycobacteri
31	231	10.6	436	1 THCB_RHOER	P43492 rhodococcus
32	229	10.5	402	1 Y76E_MYCTU	P77902 mycobacteri
33	218	10.0	422	1 CPXC_AGRTE	P24466 agrobacteri

## ALIGNMENTS

RESULT	1	STANDARD	PRT	414 AA.
CPXA_PSEPU				
ID CPXA_PSEPU				
AC P00183:				
DT 21-JUL-1986 (Rel. 01, Created)				
DT 13-AUG-1987 (Rel. 05, Last sequence update)				
DT 15-FEB-2000 (Rel. 39, Last annotation update)				
DE CYTOCHROME P450-CAM (EC 1.14.15.1) (CAMPOR 5-MONOOXYGENASE).				
GN CAMC OR CYP101.				
OS Pseudomonas putida.				
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;				
OC Pseudomonas.				
RN [1]				
RP SEQUENCE FROM N.A.				
RX MEDLINE: 86111751.				
RA Unger B.P., Gunsalus I.C., Sliker S.G.;				
RT "Nucleotide sequence of the Pseudomonas putida cytochrome P-450cam				
RT gene and its expression in Escherichia coli.";				
RT J. Biol. Chem. 261:1158-1163(1986).				
RN [2]				
RP SEQUENCE OF 385-414 FROM N.A.				
RC STRAIN-ATCC 17453;				
RX MEDLINE: 90130389.				
RA Koga H., Yamaguchi E., Matsunaga K., Aramaki H., Horiuchi T.;				
RT "Cloning and nucleotide sequences of NADH-putidaredoxin reductase				
RT gene (camA) and putidaredoxin gene (camB) involved in cytochrome				
RT P-450cam hydroxylase of Pseudomonas putida.";				
RT J. Biochem. 106:831-836(1989).				
RL [3]				
RN SEQUENCE.				
RP MEDLINE: 83030788.				
RA Hanlu M., Armes L.G., Yasunobu K.T., Shastri B.A., Gunsalus I.C.;				
RT "Amino acid sequence of the Pseudomonas putida cytochrome P-450. II.				
RT Cyanogen bromide peptides, acid cleavage peptides, and the complete				
RT sequence.";				
RT J. Biol. Chem. 257:12664-12671(1982).				
RL [4]				
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).				
RX MEDLINE: 86143817.				
RA Finzel B.C., Weber P.C., Hardman K.D., Salame F.R.;				
RT "Structure of ferricytochrome c' from Rhodospirillum rubrum at				
RT 1.67-A resolution.";				
RT J. Mol. Biol. 186:627-643(1985).				
RN [5]				
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).				
RX MEDLINE: 86059514.				
RA Poulos T.L., Finzel B.C., Gunsalus I.C., Wagner G.C., Kraut J.;				
RT "The 2.6-A crystal structure of Pseudomonas putida cytochrome P-450.";				
RT J. Biol. Chem. 260:16122-16130(1985).				
RN [6]				
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).				
RX MEDLINE: 98019009.				
RA Schlichting I., Jung C., Schulze H.;				
RT "Crystal structure of cytochrome P-450cam complexed with the (1S)-				
RT camphor enantiomer.";				
RT FEBS Lett. 415:253-257(1997).				

RN [7]  
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).  
RX MEDLINE: 98313255.  
RA Vidakovic M., Sligar S.G., Li R., Poulos T.L.;  
RT "Understanding the role of the essential Asp251 in cytochrome P450cam  
RT using site-directed mutagenesis, crystallography, and kinetic solvent  
RT isotope effect.";   
RL Biochemistry 37:9211-9219(1998).  
RN [8]  
RP STRUCTURE BY NMR.  
RX MEDLINE: 97459726.  
RA Moura C., Bondon A., Simmoneaux G., Jung C.;  
RT "1H-NMR study of diamagnetic cytochrome P450cam: assignment of heme  
RT resonances and substrate dependence of one cysteine beta proton.";   
RL FEBS Lett. 414:203-208(1997).  
CC -1- FUNCTION: INVOLVED IN A CAMPHOR OXIDATION SYSTEM.  
CC -1- CATALYTIC ACTIVITY: CAMPHOR + PUTIDAREDOXIN + O(2) = 5-EXO-  
CC HYDROXYCAMPHOR + OXIDIZED PUTIDAREDOXIN + H(2)O.  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M12546; AAA25760.1; -;  
CC EMBL: D00528; BAA00412.1; -;  
DR PIR; A25660; Q4PSCP.  
DR PDB; 2CPE; 15-APR-91.  
DR PDB; 3CPE; 15-APR-91.  
DR PDB; 4CPE; 15-JUL-91.  
DR PDB; 5CPE; 15-JUL-91.  
DR PDB; 6CPE; 15-JUL-91.  
DR PDB; 7CPE; 15-JUL-91.  
DR PDB; 8CPE; 15-JUL-91.  
DR PDB; 1CP4; 15-JAN-93.  
DR PDB; 2CP4; 15-JAN-93.  
DR PDB; 3CP4; 15-JAN-93.  
DR PDB; 4CP4; 15-JAN-93.  
DR PDB; 5CP4; 16-SEP-98.  
DR PDB; 6CP4; 16-SEP-98.  
DR PDB; 1N0O; 08-MAR-96.  
DR PDB; 1PHA; 31-OCT-93.  
DR PDB; 1PHB; 31-OCT-93.  
DR PDB; 1PHC; 31-OCT-93.  
DR PDB; 1PHD; 31-OCT-93.  
DR PDB; 1PHE; 31-OCT-93.  
DR PDB; 1PHF; 31-OCT-93.  
DR PDB; 1PHG; 31-OCT-93.  
DR PDB; 1AKD; 19-NOV-97.  
DR PIRAM: PFO0067; P450; 1.  
DR PROSITE: PS00086; CYTOCHROME\_P450; 1.  
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;  
KW 3D-structure.  
FT INIT\_MET 0  
FT BINDING 357 357  
FT CONFLICT 55  
FT CONFLICT 276 276  
FT CONFLICT 361 361  
FT CONFLICT 407 407  
FT TURN 16 17  
FT HELIX 20 22  
FT STRAND 23 23  
FT TURN 28 29  
FT TURN 32 33  
FT HELIX 34 36  
FT HELIX 38 46  
FT TURN 48 49  
FT STRAND 53 56  
FT HELIX 58 60

FT	STRAND	62	65
FT	HELIX	78	76
FT	TURN	68	80
FT	STRAND	81	82
FT	TURN	83	84
FT	HELIX	90	95
FT	TURN	99	102
FT	TURN	105	107
FT	HELIX	108	119
FT	HELIX	121	145
FT	TURN	146	146
FT	STRAND	147	149
FT	HELIX	150	153
FT	TURN	154	156
FT	HELIX	157	167
FT	TURN	168	168
FT	HELIX	171	185
FT	HELIX	193	213
FT	TURN	219	224
FT	TURN	225	225
FT	STRAND	227	228
FT	TURN	229	230
FT	STRAND	231	232
FT	HELIX	235	250
FT	TURN	251	252
FT	HELIX	253	266
FT	HELIX	268	276
FT	HELIX	278	291
FT	TURN	292	292
FT	STRAND	295	295
FT	STRAND	297	301
FT	TURN	305	307
FT	TURN	308	309
FT	STRAND	310	312
FT	TURN	314	315
FT	STRAND	317	320
FT	HELIX	322	324
FT	TURN	325	327
FT	TURN	329	331
FT	TURN	335	336
FT	TURN	340	341
FT	TURN	349	350
FT	HELIX	353	355
FT	TURN	358	359
FT	HELIX	360	377
FT	STRAND	382	383
FT	TURN	385	386
FT	STRAND	391	392
FT	STRAND	396	396
FT	STRAND	398	399
FT	STRAND	403	405
FT	HELIX	408	410
SO	SEQUENCE	414 AA;	46538 MW; ECA610293A9D6207 CRC64;

Query Match 99.5%; Score 2169; DB 1; Length 414;  
Best Local Similarity 99.5%; Pred. No. 2.6e-158;  
Matches 412; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY	1	TTETIOSNANLAPLPHVPEHLVFDMDYVPSNLSAGVQEMAVLQBSNVPDLVWTRCNG	60
DB	1	TTETIOSNANLAPLPHVPEHLVFDMDYVPSNLSAGVQEMAVLQBSNVPDLVWTRCNG	60
OY	61	GHMIATRGOLIRAYEDYRHSSECPPIPEAGEAYDFITSMDDPEQORFALANQVYG	120
DB	61	GHMIATRGOLIRAYEDYRHSSECPPIPEAGEAYDFITSMDDPEQORFALANQVYG	120
OY	121	MPVVDKLENNIQELACSLIESLRPOGQCNFTEDYAEFPPIRIMLLAGLPEEDIPIHLKYL	180
DB	121	MPVVDKLENNIQELACSLIESLRPOGQCNFTEDYAEFPPIRIMLLAGLPEEDIPIHLKYL	180
OY	181	TDQMTRPDGSMTFAEAKKALYDYLPIIEQROKPGTDAISIVANGQVNGRPITSEAKR	240

Db	181	TDQMRRPQSGMSTFAAKENALDYLLIPITIEQRQRKPGCTAISIVANGQVNGRPTSDCAKR	240
QY	241	MGCLLVAGGLDTVAVNFSLFSMEFLAKSPENHRELIERELLIPACEELLRRSLVADGRI	300
Db	241	MGCLLVAGGLDTVAVNFSLFSMEFLAKSPENHRELIERERIPAAEEELLRRSLVADGRI	300
QY	301	LTSDFEFHGVOLKCKDQDILLPOMLSGLDERKNACPMHNDPESROKSHHTFGHSHLCIQ	360
Db	301	LTSDFEFHGVOLKCKDQDILLPOMLSGLDERKNACPMHNDPESROKSHHTFGHSHLCIQ	360
QY	361	HLAREIIVTLKEMWLTIRIPDPSIAPGAQIQHKSIVSGVALPLVMDPATTKAV	414
Db	361	HLAREIIVTLKEMWLTIRIPDPSIAPGAQIQHKSIVSGVALPLVMDPATTKAV	414
RESULT 2			
Y4VG_RHISN	ID	Y4VG_RHISN	STANDARD; PRT: 414 AA.
AC	O53215;		
DT	01-NOV-1997 (Rel. 35, Created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)		
DT	15-DEC-1998 (Rel. 37, Last annotation update)		
DE	PROBABLE CYTOCHROME P450 Y4VG (EC 1.14.14.-).		
GN	Y4VG.		
OS	Rhizobium sp. (strain NGR234).		
OC	Plasmid sym PNGR234e.		
OC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;		
OC	Rhizobiaceae; Rhizobium.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE: 97305956.		
RA	Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,		
RA	Petret X.:		
RT	"Molecular basis of symbiosis between Rhizobium and legumes.";		
RL	Nature 387:394-401(1997).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE: 963899014.		
RA	Freiberg C., Petret X., Broughton W.J., Rosenthal A.;		
RT	"Sequencing the 500-kb GC-rich symbiotic replicon of Rhizobium sp.		
RT	NGR234 using dye terminators and a thermostable 'sequenase': a		
RT	beginning.";		
RL	Genome Res. 6:590-600(1996).		
CC	-1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE		
CC	MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED		
CC	COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.		
CC	-1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> ).		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> .		
CC	-----		
DR	EMBL: Z68Z03: CAAG2422.1: -		
DR	EMBL: AE000101: AAB91895.1: -		
DR	HSSP: P00183: 3CP4.		
DR	PFAM: PF00067: P450. 2.		
DR	PROSITE: PS00086: CYTOCHROME_P450; FALSE NEG.		
KW	Hypothetical protein: Oxidoreductase; Monooxygenase;		
KW	Electron transport; Heme; Plasmid.		
FT	BINDING 364 364 HEME (BY SIMILARITY).		
SQ	SEQUENCE 414 AA: 45810 MW: 888598E99315BB84 CXC64;		

Query Match	26.8%	Score 584.5	DB 1	Length 414
Best Local Similarity	33.8%	Pred. No. 1.9e-37		
Matches 141	Conservative 65	Mismatches 172	Indels 39	Gaps 9
QY	13 PLPLPVPPEHLVDFEDMYNPSNLSA---GVQGEAMVNLQESNVDFLWY----TRCNGHMTA 65			
	1:1 111 11 1:1 ::	1:1 1:1 ::	11 1:1	

```

Db 13 PIPHVPALVRHHSLSFTSPGMAPTPNDPDHAANAACVADDOCPRIFFSPSNRDRGTVI 72
Qy 66 TRGQLIREAYDYRRHFSESCPIPREACEAYDIPITSDMPEDQOFALANOVGMYPVD 125
Db 73 TRARQORFVLEDTETFESSHRSIFASALCEHMPVIPLELDPAHGVALNLPLFFSSRYL 132
Qy 126 KLENNIOELASLIESL-RPGGOCNFETDVAEPPIRIEMLAGPE-----BDI 174
Db 133 ALEPTIHARACALIDCIKAKETSCVMDFALPFEVSFSLGISOBSREVLVGMWSDL 192
Qy 175 PHLYTUDQMTRPDQSMFPAEAKELALYDLPITIEQ---BRQPGTAISIVANGVNG 230
Db 193 LH-----GN---AEKRAAASVAFIDEMAAKRSKSAVDMEFFVYQAKIEG 237
Qy 231 RPIITSEAKRMCGLLVGLDVTYVNFISFSMEFLAKSPENROELIERPELLPAACELLR 290
Db 238 RSLTVEEVRGIGVLEFVAGLDVTAALIGFDMAVLAIRNPKHOLLRNBPARGLAEEILLR 297
Qy 291 RPSLVADQRILTSDFEFGVQOLKKGDDLLPQMLSGLDERKNACPMHNDSEKOVSHTF 350
Db 298 AVSTVQIIRVATKDIIEFGVPIREDVYSCPMIANRPSSEKCPNTDLARODQHTAF 357
Qy 351 GGHSGILGOLHAREILVTPLKEMTLRIPDPSIAPG-RQIOHKSIGVGYQALPLW 406
Db 358 GYCPHLCGAILARREIYIGLEMLARIPARIRIKETGAPITH-GGHVFGISNILLTW 413

RESULT 3
CPXM_BACME STANDARD: PRT: 410 AA.
AC 006069;
ID CPXM_BACME STANDARD: PRT: 410 AA.
AC 006069;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE CYCROCHROME P450(MEG) (EC 1.14.99.-) (STEROID 15-BETA-HYDROXYLASE)
DE (STEROID 15-BETA-MONOOXYGENASE).
GN CYP106A2.
OS Bacillus megaterium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/staphylococcus group; Bacillus.
RN [1]
RA SEQUENCE FROM N.A.
RC STRAIN-ATCC 13368;
RX MEDLINE; 94049677.
RA Rauschenbach R., Isernhagen M., Noeske-Jungblut C., Boidol W.,
RA Siewert G.;
RT "Cloning, sequencing and expression of the gene for cytochrome
RT P450meg, the steroid-15 beta-monooxygenase from Bacillus megaterium
RT ATCC 13368."
RL Mol. Gen. Genet. 241:170-176(1993).
RN [2]
RA CHARACTERIZATION.
RC STRAIN-ATCC 13368;
RX MEDLINE; 79194051.
RA Berg A., Ingelman-Sundberg M., Gustafsson M.;
RT "Purification and characterization of cytochrome P-450meg.";
RL J. Biol. Chem. 254:5264-5271(1979).
RN [3]
RA CHARACTERIZATION.
RC STRAIN-ATCC 13368;
RX MEDLINE; 80201079.
RA Berg A., Rafter J.J.;
RT "Studies on the substrate specificity and inducibility of cytochrome
RT P-450meg.";
RL Biochem. J. 196:781-786(1981).
CC -1- FUNCTION: HAS THE CAPACITY TO HYDROXYLATE CERTAIN STEROIDS IN THE
CC -1- BETA POSITION.
CC -1- SIMILARITY: BELONGS TO THE CYCROCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```



RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AF015825; AAC46317.1; -.  
DR EMBL: Z99110; CAB13078.1; -.  
DR SUBMITTER: BG13195; YUIB.  
DR PFAM: PF00067; P450; 1.  
DR PRINTS: PR00359; BP450.  
DR PROSITE: PS00086; CYTOCHROME\_P450; 1.  
KW Hypothetical protein: Oxidoreductase; Monooxygenase; Heme.  
FT BINDING 349 349 HEME (BY SIMILARITY).  
SQ SEQUENCE 396 AA; 44990 MW; 9A89CF12613DBCFB CRC64;  
  
Query Match 17.1%; Score 372; DB 1; Length 396;  
Best Local Similarity 30.7%; Pred. No. 3e-21;  
Matches 101; Conservative 54; Mismatches 156; Indels 18; Gaps 7;  
  
OY 71 IREAYEDYRHSSRCPPFIPRAGEAYDFIPMSDPPQORORALANOVGVYDKLENR 130  
Db 56 VKKVGKDELFS-C-MPOQTSSIGNSI-IMDDPKHTKIRSVYVKKFTYRVMKQWEPK 111  
OY 131 IOELACSLIESLRQGCNFTEDYAEPPIRIFMLAGLPREDIPHLKYLTDOMT----- 185  
Db 112 IOETDELIOKFGQSEEDLVHDSYPLPVIVISELIGVPSAHMQGFAMWDLVSTPKD 171  
OY 186 -RPGSMFPAE-----AKALYDYLPIIEQRKQKPTDAISIVANGVNGRPITSDKKR 240  
Db 172 KSEAEKAFLEERDKCEBELAFAGIIEEKRNRKPEODIISILVEAETGKLSGEELIP 231  
OY 241 MCGLLVAGLGVVNFSLFSMEPLAKSPENHOLELERPELIPACEBELLR-RESLVADGR 299  
Db 232 FCTLLLVAGNETTNNLSNAMYSLIETPGVYELRSHELMPOAVEELRRAPAPVLR 291  
OY 300 ILTSDYEFHGVOLKGDQILLPOMLSGLDERKNACPMHVDERSOKVSHTTGHSGLCLG 359  
Db 292 IAKRDTIELGHLIKEGDVAFLFVASANDKAKFDRPHMFDLRHNPRIATGCHGHPCLG 351  
OY 360 QHLAREIIVTLKEMLRIPDF---SLAP 385  
Db 352 APLARLEANIALTSLISAFPMHECVSITP 380  
  
RESULT 6  
Y218\_MYCTU STANDARD: PRT; 398 AA.  
AC 053563;  
DT 15-FEB-2000 (Rel. 39, Created)  
DT 15-FEB-2000 (Rel. 39, Last sequence update)  
DE PUTATIVE CYTOCHROME P450 RV3518C (EC 1.14.-.-).  
GN RV3518C OR MTW023.25C.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H37RV;  
RC MEDLINE: 98295987.  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holtroyd S.,  
RA Hornsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,

RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,  
RA Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence."  
RL Nature 393:537-544(1998).  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AL022022; CAA11755.1; -.  
DR TUBERCULIST; RV3518C; -.  
DR PFAM: PF00067; P450; 1.  
DR PRINTS: PR00359; BP450.  
DR PROSITE: PS00086; CYTOCHROME\_P450; 1.  
KW Hypothetical protein: Oxidoreductase; Monooxygenase; Heme.  
FT BINDING 340 340 HEME (BY SIMILARITY).  
SQ SEQUENCE 398 AA; 44398 MW; BCF3C23BCB5767F CRC64;  
  
Query Match 16.8%; Score 366; DB 1; Length 398;  
Best Local Similarity 28.5%; Pred. No. 8.6e-21;  
Matches 108; Conservative 69; Mismatches 168; Indels 34; Gaps 8;  
  
OY 55 WTRCN-----GGHMTATRGQLIREAYEDYRHSS-----ECPPFIPRAGEAYDFIPT 101  
Db 26 WMRAQPVFDRNRGLAASTYQAVIDAERPELSNAGIRPDQALP-----MMI 76  
OY 102 SMDDPEQORORALANOVGVYDKLENRIOELACSLIESLRQGCNFTEDYAEPPIR 161  
Db 77 DMDDPAHLRRLKRLVAGTRRRKRVKDEKSIALCDTLIDAVCEKCEQDFVLDLAPLEMA 136  
OY 162 IFMLLAGL-PEEDIPLHLK-----YLFDMTRPDGSMTPFAKALVDYLPIIEORQ 213  
Db 137 VIGDMGVPRQDRMFLMSDDLVTFLSSHWQSDPOTTM-DARAAYUDFTRATIAARRA 195  
OY 214 KPGTDAISIVANGVNGRPITSDKAKRCGLLVGGLDVTYVNFSLFSMEPLAKSPENH 273  
Db 196 DPTDOLVSLVSEYDGRSLSDDELVMETLLILGQDETTHHTLSGTEQLLRNDQMDL 255  
OY 274 LIERPELIPACEBELLRRESLVAD-GRILTSDYEFHGVOLKGDQILLPOMLSGLDERKN 332  
Db 256 LQRPDSLIPGAIEEMLRMTAVKMKRCVLTADTEFHGALCAGKMKMLDFESANFDEAVF 315  
OY 333 ACPMHVDSRQKVSHTTGHGSHCLIGHLLARREIIVTLKEMLRIPDFSIAPGQI--Q 390  
Db 316 CEPKEDVQRRPNSHLAGFTHTFCLGNQLARLELSTMTREVLRLPLRVADSDVPL 375  
OY 391 HKSGIVSGVALPLWMDPA 409  
Db 376 RPAENVSGLESMPTVFTPS 394  
  
RESULT 7  
CPXM\_BACSU STANDARD: PRT; 405 AA.  
ID CPXM\_BACSU  
AC P27632;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 15-FEB-2000 (Rel. 39, Last annotation update)  
DE CYTOCHROME P450 109 (EC 1.14.-.-) (ORF405).  
GN CYP109.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-W23;





```
AC OS0696;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 15-FEB-2000 (Rel. 39, Last annotation update)
DE PUTATIVE CYTOCHROME P450 RV2266 (EC 1.14.-.-).
GN RV2266 OR MTCY339.44C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE: 98295987.
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Sultson J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z77163; CAB00969.1; -.
DR HSSP: P33006; ICPT.
DR TUBERCULIST: RV2266; -.
DR PFAM: PF00067; P450; 1.
DR PRINTS: PR00359; BP450.
DR PROSITE: PS00086; CYTOCHROME_P450; FALSE_NEG.
KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.
FT BINDING 379 379 HEME (by SIMILARITY).
SQ SEQUENCE 428 AA; 47824 MW; 76B1F3C5AE348591 CRC64;

Query Match 15.7%; Score 342.5; DB 1; Length 428;
Best Local Similarity 26.3%; Pred. No. 5.9e-19;
Matches 118; Conservative 71; Mismatches 188; Indels 69; Gaps 17;

5 IOSNANLAPRPYVPEHVLVDFEDMWNPSNLSAGVOEAMAVLOE-----SNVPLV 54
Db 7 IATRVNGRP-PPEVP---IADIEL-----GSLDFWALDDVDVDFGAFATLRAPISF 54
OY 55 WTR-----CNGHMWATRCQLREAYEDY-----RH-----ECPFIPREA 92
Db 55 WPTIELPGFVAGNGHMAITK-----YDVEFYASRHDPIDSSYPNITINDOTPELAYEF 107
OY 93 GEATDFITSDPPEQRFRLANOVGMFVYDKLENIQELACLISL---RPGGCN 149
Db 108 GSM-----IVLDDPRHQLRSIVSRATPKVVARIEAIVRRRAHLVSSMANNDRQAD 162
OY 150 FTEVDAEPFPIRIFMLLAGPEEDIPHLKYLTQMT---RPGSMTFPEAKEALYD---Y 203
Db 163 LVSELAGLPLQIICDMNGIFKADHQRIFFHNTYNILGFGDDPLADLDFEPFQVSDIDTAY 222
OY 204 LIPITQRQRKPGTDAISIVANGVNGRPITSDEAKRMCGLLVGLGDTVYNFLSFSMEF 263
Db 223 ATALAEDRVNHHDLTSLVEAEVDGERSLSREIASFFILLVAGNETRNAITHGVIA 282
OY 264 LAKSEHNOF-LIERPELIPACRELLARFS-LVADGRILTSVDFEFGVQKKGQIILP 321
Db 283 LSRPEQDRMWSDFGLAPRAVEIIVWASPVVYMRRTLTQDIELRGTMAAGDKVSLW 342
```

```
OY 322 QMLGLDERKNACPMHVDNFSKQVSHTFE-GGSHLCLGQHLARREITVLEKWLTRIPD 380
Db 343 YCSANROESKFRADPMTFPLARNPNPHLFGGGGAHFCGANLARREIRVAPDELRQMPD 402
OY 381 FSIAPGAOIQRKSGIVSGVALPLVW 406
Db 403 V-VATEEPARLLSQFIRHGIKTLPVYW 427

RESULT 10
CPXL_BACME STANDARD; PRT; 410 AA.
ID CPXL_BACME
AC P14762;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE CYTOCHROME P450(BM-1) (EC 1.14.14.1).
GN CYP106.
OS Bacillus megaterium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-25.
RC STRAIN-ATCC 14581;
RX MEDLINE: 90089408.
RA He J.S., Ruettinger R.T., Liu H.-M., Fulco A.J.;
RT "Molecular cloning, coding nucleotides and the deduced amino acid
RT sequence of P-450BM-1 from Bacillus megaterium."
RL Biochim. Biophys. Acta 1009:301-303(1989).
RN [2]
RP SEQUENCE OF 1-25 FROM N.A.
RX MEDLINE: 95355495.
RA He J.S., Liang Q., Fulco A.J.;
RT "The molecular cloning and characterization of BM1P1 and BM1P2
RT proteins, putative positive transcription factors involved in
RT barbiturate-mediated induction of the genes encoding cytochrome
RT P450BM-1 of Bacillus megaterium."
RL J. Biol. Chem. 270:18615-18625(1995).
CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
CC MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED
CC COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X16610; CA34612.1; -.
DR EMBL: S79230; AAC60495.1; -.
DR PIR: S07764; OAB56M.
DR HSSP: P33006; ICPT.
DR PFAM: PF00067; P450; 1.
DR PRINTS: PR00359; BP450.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.
FT BINDING 356 356 HEME.
SQ SEQUENCE 410 AA; 47460 MW; C9AE293E76745387 CRC64;

Query Match 15.7%; Score 341.5; DB 1; Length 410;
Best Local Similarity 28.0%; Pred. No. 6.6e-19;
Matches 97; Conservative 57; Mismatches 169; Indels 23; Gaps 7;

63 WIATGQILREAYEDYRHFSSSE---CPPIPRAGGAYDFIP-----TSMDPPEQRFRLA 115
Db 45 WNVFQYENKQVLSNTRDFSSDGGRTTFVGDNSKKKTSPTTNLTNIDPDDHKRAKSL 104
OY 116 NOVGMFVYDKLENIQELACLISLRPGOCNFTEDYAEFPPIRIFMLLAGPEEDIP 175
```

```

Db      105  AAFAFPRLSKNNEPRITKQIAADLVAEIAQKNSTINIVDDLSPPRLVIADLFEVPPKDRX 164
QY      176  HLKYLTDQMTRPDGSMTFAEAKE-----ALYDLIPIIEOROKPGDAISIVANGQY 228
Db      165  QPKKWWDLIFOPYDERLEIEIEDEKORAGAEFQYLYPIVEIKRSMJLSDIITDLOAEV 224
QY      229  NGRPTISDAKMMCGLLIYGGIDTVYNFL-SYSMELASPEIKROLIERPELIPACEE 287
Db      225  DEETFTDEETVIAHYMTLLLAGVETTSHTAIAANPEYSLYDKSLYSFLRNRLNRLAPAAVE 284
QY      288  LLR-FRSLVADNRILITSDVEFPGVOLKSKDQILLPOMLSGLDERKNKACPMHYDFSR-QKY 345
Db      285  MLRYRHRHSIRBRDTYKQDNELLGVKTKKDDVYIAMSACNMDETMEFNFSVDIHPTUK 344
QY      346  SHTTGCHGSHLGLQHLARREIIVTLKELTR-----IPDSIAP 385
Db      345  KHLITGNGPFCIAGPLARLEFMKILLEAFLEAFSHIEPFEDELEP 390

RESULT 11
CPXK_SACER STANDARD; PRT; 405 AA.
AC P133271;
DT 01-FEB-1994 (Rel. 28, Created)
DI 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CYTOCHROME P450 107B1 (EC 1.14.-.-) (P450C11B1).
GN CYP107B1.
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Pseudonocardiaceae; Pseudonocardiaceae;
OC Saccharopolyspora.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 172-203 AND 302-321.
RC STRAIN-NRRL 2338;
RX MEDLINE: 92121109.
RA Andersen J.F., Hutchinson C.R.;
RT "Characterization of Saccharopolyspora erythraea cytochrome P-450
RT genes and enzymes, including 6-deoxyerythronolide B hydroxylase.";
RL J. Bacteriol. 174:725-735(1992).
CC 1- FUNCTION: NOT KNOWN, PROBABLY INVOLVED IN THE CATABOLISM OF
CC OCANE AND GAIAICOL. IT DISPLAYS A WEAK ACTIVITY IN THE O-
CC DEMETHYLATION OF 7-ETHOXYCOCAIN.
CC 1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M83110; AAA26483.1; -.
CC PIR: B42606; B42606.
DR HSSP: Q00441; 10XA.
DR PEARL: PF00067; P450. 1.
DR PRINTS: PR00359; BP450.
DR PROSITE: PS00086; CYTOCHROME_P450. 1.
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.
FT BINDING 352 352 HEME (BY SIMILARITY).
FT SEQUENCE 405 AA: 45238 MW; 71C93CEC1FDC53FD CRC64;

Query Match 15.6%; Score 339.5; DB 1; Length 405;
Best Local Similarity 25.8%; Pred. No. 9.2e-19;
Matches 108; Conservative 74; Mismatches 183; Indels 53; Gaps 10.

QY 18 VPEHLVFP-----FDMYNSNSLSAGVOAEAMVAYENVPLVWTRGNG-GHWIATRKQLIR 72
Db 6 VPDLLAFPDFAQODKHN-----RRARRREPPYQRI-RTYNGDGLAWLITREYDK 53
QY 73 EAYEDYRHFSSECPPIPREAGEAYDFIPTSM-----DPEEORQER 112

```

[illegible]





DR PROSITE; PS00086; CYTOCHROME\_P450.1.  
KW Biotin biosynthesis; Oxidoreductase; Monooxygenase;  
KW Electron transport; Heme.  
FT BINDING 345 HEME (BY SIMILARITY).  
SQ SEQUENCE 395 AA; 44865 MW; EAC3AF2637ACE1A C9C64;

Query Match	14.8%	Score 322.5;	DB 1	Length 395;
Best Local Similarity	27.2%	Pred. No. 1.8e-17;		
Matches 88; Conservative	64;	Mismatches 138;	Indels 33;	Gaps 7;

Qy	75	YEDRNHSESCEPFLIPREGEVAYDIEPTSMRPEOROSRILANOVAMPYDKLENRJOEL	134
Db	70	YODLSHONQNMILFONQ-----PDRNRILATLSAGFAFPRTTESQVPIIET	115
Qy	135	ACSLIESIRPGOGCNFIEDVAYERRPIKEML--LAGPREDIPHLKTLTPOIMRPGCSMT	192
Db	116	VHNLIDVOGKKMKMEVISDEFA--FPLASFVYANLIGVEEDREOLKEMASLJO--TID	170
Qy	193	FAEAKREALYD-----YLPIEIORROKPGTDALISIVANGOVGRPITSDCAKRMCS	242
Db	171	FTRSRKATTEGNIAMAYOMAFYFKELQKKRKNRQDOMISMILKREKDK--LTBEAASVC	229
Qy	243	GLLLVGGGLDYVNFPLSMTSMERLASPREHROELIRREPLIRACSELLRRS--LYADGRIL	301
Db	230	ILLIAGHETTVNLIINSVLCILLOHNPOLKLRBNPRLIGTAVEECIRYSPOTMARA	289
Qy	302	TSDVEFNQVOKKQDNLIPOMLSGLDERKACSMYNDVFSROKXSHTFENGSHILSGON	361
Db	290	SEDDIDCGVTLIRGEBOYULLIGGANNRDPSTPTNDRVDVDIRSPRHLISFGNHNHVGSS	349
Qy	362	LARELIYTLKEMLTIRPDPSIA	384
Db	350	LARLEAOIAINTLLOMRPSLMA	372

Search completed: October 4, 2000, 13:04:53  
Job time: 1684 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 12:39:36 ; Search time 79.83 Seconds  
(without alignments)  
359.568 Million cell updates/sec

Title: US-09-246-451-12  
Perfect score: 2179  
Sequence: 1 TTTTQSNANLAPLPHPVPE.....IVSGVOALPLVWDPATTRAV 414

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP\_TREMBL\_12:\*  
2: SP\_archaea:\*  
3: SP\_bacteria:\*  
4: SP\_fungi:\*  
5: SP\_human:\*  
6: SP\_invertebrate:\*  
7: SP\_mammal:\*  
8: SP\_mhcc:\*  
9: SP\_mhcc:\*  
10: SP\_mhcc:\*  
11: SP\_mhcc:\*  
12: SP\_mhcc:\*  
13: SP\_mhcc:\*  
14: SP\_mhcc:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	371.5	17.0	416	2	087605 streptomyc
2	370	17.0	397	2	059523 micromonos
3	360	16.5	396	2	052544 amycolatops
4	353	16.2	410	2	09X5P9 streptomyc
5	352.5	16.2	404	2	054302 streptomyc
6	350.5	16.1	407	2	059819 streptomyc
7	332	15.2	406	2	059723 pseudomonas
8	331.5	15.2	420	2	09ZH01 streptomyc
9	327.5	15.0	376	2	031785 bacillus su
10	327	15.0	407	2	09X5P8 streptomyc
11	322	15.0	411	2	032460 actinobacteri
12	322	14.8	410	2	09X9P7 streptomyc
13	319.5	14.7	410	2	059831 streptomyc
14	318	14.6	400	2	09XDB0 mycobacteri
15	311	14.3	310	2	032927 mycobacteri
16	310	14.2	406	2	087192 streptomyc
17	309.5	14.2	397	2	024727 nocardioide
18	309	14.2	417	2	059910 streptomyc
19	307	14.1	406	2	085697 streptomyc

20	297	13.6	388	2	P77977 streptomyc
21	290.5	13.3	337	2	052816 amycolatops
22	288.5	13.2	406	2	087675 amycolatops
23	286	13.1	351	2	052572 amycolatops
24	285	13.1	421	2	052561 amycolatops
25	273.5	12.6	310	2	P96562 amycolatops
26	270	12.4	403	2	052560 amycolatops
27	268.5	12.3	411	2	09X803 streptomyc
28	267.5	12.3	386	2	059921 streptomyc
29	266	12.2	398	2	087674 amycolatops
30	265	12.2	395	2	09X5T2 streptomyc
31	258.5	11.9	437	2	09ZFC0 mycobacteri
32	255	11.7	411	2	060005 s putative
33	254.5	11.7	511	2	052569 amycolatops
34	254	11.7	406	2	052823 amycolatops
35	249	11.4	426	2	09ZFC3 mycobacteri
36	233	10.7	433	2	09XC6 streptomyc
37	231	10.6	398	2	059079 amycolata a
38	226	10.4	391	2	052822 amycolatops
39	224.5	10.3	391	2	087673 amycolatops
40	213	9.8	397	2	052802 amycolatops
41	201.5	9.2	419	2	050242 agrobacteri
42	188.5	8.7	119	2	085655 streptomyc
43	185	8.5	120	2	085653 streptococ
44	180.5	8.3	313	2	09X418 myxococcus
45	170	7.8	120	2	085650 streptomyc

## ALIGNMENTS

RESULT 1  
ID 087605 PRELIMINARY: PRT: 416 AA.  
AC 087605;  
DT 01-NOV-1998 (TREMBL) 08, Created)  
DT 01-NOV-1998 (TREMBL) 08, Last sequence update)  
DT 01-NOV-1999 (TREMBL) 12, Last annotation update)  
DE CYTOCHROME P450 MONOOXYGENASE.  
GN PICK OR PICK.  
OS Streptomyces violaceus (Streptomyces venezuelae), and  
OS Streptomyces venezuelae.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC15439;  
RA BETLACH M.C., KEALEY J.T., ASHLEY G.W., MCDANIEL R.;  
RT "Characterization of the macrolide P450 hydroxylase from Streptomyces  
RT venezuelae which converts narbomycin to plectromycin.";  
RL Biochemistry 0:0-0(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES-S.VENEZUELA; STRAIN-ATCC15439;  
RA MEDLINE: 99051447.  
RX XUE Y., WILSON D., ZHAO L., LIU H.-W., SHERMAN D.H.;  
RT "Hydroxylation of macrolactones XC-17 and narbomycin is mediated by  
RT the PKC-encoded cytochrome P450 in Streptomyces venezuelae.";  
RL Chem. Biol. 5:661-669(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES-S.VENEZUELA; STRAIN-ATCC15439;  
RX MEDLINE: 98445333.  
RA XUE Y., ZHAO L., LIU H.-W., SHERMAN D.H.;  
RT "A gene cluster for macrolide antibiotic biosynthesis in Streptomyces  
RT venezuelae: architecture of metabolic diversity.";  
RL Proc. Natl. Acad. Sci. U.S.A. 95:12111-12116(1998).  
DR EMBL: AF079139; AAC68886.1; -  
DR HSBP: Q00441; 10XA.  
DR PFAM: PF00067; P450; 1.  
DR PRINTS: PR00359; BP450.  
KW Monooxygenase.





QY 188 DGSMTFAEAKALYDYLPIIEQRORCKPTDAISIVANGVNGRPTISDEAKRMGCLLV 247  
DB 182 AEA---AAAEVAGKFEFDEVIERRRQRDDLSLV-----EDLPQEBELRNITLTLF 233  
QY 248 GGLDTVWFVLFSEMEFLAKSPENROELIERPELLPACCELLRPSL--VADGRILSDY 305  
DB 224 AGYETTEGALATGVFALHHTDOLAALRAPEPKIDAAIEELLRYLTVMQYHTYRTALRBY 293  
QY 306 EFHGVLAKKGGQIILLPOMLSGLDERKNACPMHVDERSCKVSHTEFGHSHCLIGOHLLAR 365  
DB 294 KLEBELIKKGGTIVTSLPAARNRDPKRCGPALDIERTSGHVAFGFISHOCLGONLARI 353  
QY 366 EIIYTLKEMLTRIPDFESIA-PGAQIOHK-SCIVSGVALPLVM 406  
DB 354 ELRAGFTALLRAPPFLRLAVPADEVPRLKSGVSVKLLPYSW 396

RESULT 4  
Q9X5P9 PRELIMINARY; PRT: 410 AA.  
ID 09X5P9  
AC 09X5P9  
DT 01-NOV-1999 (TREMblrel. 12, Created)  
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)  
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)  
DE CYTOCHROME P450 HYDROXYLASE ORF3.  
OS Streptomyces lavendulae.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NRRL 2564.  
RX MEDLINE: 99201491.  
RA MAO Y.Q., VAROGU M., SHERMAN D.H.;  
RT "Molecular characterization and analysis of the biosynthetic gene  
RT cluster for the antitumor antibiotic mitomycin C from Streptomyces  
RT lavendulae NRRL 2564."  
RL Chem. Biol. 6:251-263(1999).  
DR EMBL: AF127374; AAD28449.1; -  
SO SEQUENCE 410 AA; 45887 MW; 507C7F38 CRC32;

Query Match 16.2%; Score 353; DB 2; Length 410;  
Best Local Similarity 27.2%; Pred. No. 5.6e-21;  
Matches 114; Conservative 69; Mismatches 192; Indels 44; Gaps 14;

QY 9 ANLAPLRP-HVPEHLVDFDMNPSNSAGVQEAVALQESNVDPDLVTRKNGGHMIATR 67  
DB 15 AGEAPAYPFHAPDRU--EPDPY-----WEPLRRRRLQKVTLLPYGGEAMLATR 60  
QY 68 GOLREAVEDYRHFSSE---CFPTPREAGEAYDFP-----TSMRPRQORFALAN 116  
DB 61 YODRAVAFAD-RFESROLAVARGAR-----FLPHORPPDAVLSVGGPHARLRRLVG 112  
QY 117 QVNGPVVVDKLENNRIQELACSLIESLRPOG-OCNFTEDYARPPRIRIIMLAGLREEDIP 175  
DB 113 KVFPRRAREDEMRPLIQKRADDLDAEMEMGRPADLVDEFSLPFAVSMICELLYVPRERK 172  
QY 176 HLKLTDMQTPRDSMTFAEAK---ALYDYLPIIEQRORCKPCTDAISIVANGVNGRPT 232  
DB 173 RCFVMSDALLTTT-AHPPAQVRYDMQMHDIYGLVAQRVVRPADLIGSLVTAARDEBDK 231  
QY 233 ITSDAKRMGCLLVGGLDTVWVNFLSFMEFLAKSPENROELIERPELLPACCELLRPF 292  
DB 232 LTBEELVLAAILIAGETISASQIPNFLVYLFRRPOLLEIRNDHDLIPRAVEELRFEV 291  
QY 293 SLV-VAAG--RIITSDYEFHGVLAKKGGQIILLPOMLSGLDERKNACPMHVDERSCKVSHTT 349  
DB 292 PIGYVDGPRRATEDVELGSLVAVRAGETVVSMAANDRPELFDPRDLAKRNPILG 351  
QY 350 FGHSHCLGSHLARREIYTLKEMLTRIPDFESIA-PGAQIOHKSGL-YSGVQALPLVM 406  
DB 352 FGAGPHHCLGQOLARVEIQITLTTLFRRYPRILAVPESLSWKEGILAVRGMHTMPVTM 410

RESULT 5  
ID 054302 PRELIMINARY; PRT: 404 AA.  
AC 054302  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)  
DE CYTOCHROME P450.  
GN RAFN.  
OS Streptomyces hygroscopicus.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NRRL 5491.  
RX MEDLINE: 95372374.  
RA SCHWECHE T., APARICIO J.F., MOLNAR I., KOENIG A., KHAM L.E.,  
RA HAYDOCK S.F., OLIVANK M., CAFFEY P., CORTES J., LESTER J.B.,  
RA BOEHM G.A., STAUNTON J., LEADLAY P.F.;  
RT "The biosynthetic gene cluster for the polyketide immunosuppressant  
RT rapamycin."  
RT Proc. Natl. Acad. Sci. U.S.A. 92:7839-7843(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NRRL 5491.  
RA MOLNAR I., APARICIO J.F., HAYDOCK S.F., EE KHAM L., SCHWECHE T.,  
RA KOENIG A., STAUNTON J., LEADLAY P.F., STAUNTON J., LEADLAY P.F.;  
RL Gene 0:0-0(0).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NRRL 5491.  
RA APARICIO J.F., MOLNAR I., SCHWECHE T., KOENIG A., HAYDOCK S.F.,  
RA EE KHAM L., STAUNTON J., LEADLAY P.F., LESTER J.B., BOEHM G.A.,  
RA STAUNTON J., LEADLAY P.F.;  
RL Gene 0:0-0(0).  
DR EMBL: X86780; CAA60465.1; -  
DR HSSP: Q00441; 10XA.  
DR PFM: PF00067; P450.1.  
DR PRINTS: PR00359; BP450.  
SO SEQUENCE 404 AA; 45071 MW; 05AB94DF CRC32;

Query Match 16.2%; Score 352.5; DB 2; Length 404;  
Best Local Similarity 28.9%; Pred. No. 6.1e-21;  
Matches 103; Conservative 63; Mismatches 176; Indels 15; Gaps 9;

QY 63 WIAFRGOLIREAVEDYRHFSSECFP---IPRAGEAYDF-IPTSMDPREQORFALANOV 118  
DB 50 WLAASMEVAKVFVDPFR-FSSATILGKDVPRVLAIDQVYIMLMDPEHTRLRVATKA 108  
QY 119 VGMVYDKLENNRIQELACSLIESLRPOG-OCNFTEDYARPPRIRIIMLAGLREEDIPHL 177  
DB 109 LTRSMEMALRRTOGEVADDLIDKMLAKGARADIMEDFALPPIIMICGLVPLEDQKFF 168  
QY 178 KYLDQMTPRPGSKT---FARAKENLYLPIIEQRORCKPCTDAISIVANGVNGRPT 234  
DB 169 RTWSDQML-SNGAVSQEVYMAAGSLYLLSELAEERRKQDPTNDLGLSVAROKDDRLS 227  
QY 225 SDEAKRMGCLLVGGLDTVWVNFLSFMEFLAKSPENROELIERPELLPACCELLRPSL 294  
DB 228 ETELVGFAVTLILAGYETTANAGNSVTLTLTHPEKLAELKDKSLIKAVADELRITPI 287  
QY 295 VADG---RIITSDYEFHGVLAKKGGQIILLPOMLSGLDERKNACPMHVDERSCKVSHTEFG 351  
DB 288 AKQASWMAVEDVELSGTIYKAGAAVAIQHNSANTDPKVDHPEEDIFHTSNPHMSLG 347  
QY 352 HGSILCLGSHLARREIYTLKEMLTRIPDFESIA-PGAQIOHKSGL-YSGVQALPLVM 406  
DB 348 HGAIHMGQALVRVEMOTATLSLRIPALRFVAVPEPRIKFLRGRLVPSLEALPLTW 404

```
RESULT 6
Q59819 PRELIMINARY; PRT; 407 AA.
AC Q59819:
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
DE CYTOCHROME P450 (EC 1.14.14.1).
OS Streptomyces antibioticus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95255619.
RA RODRIGUEZ A.M., OLANO C., MENDEZ C., HUTCHINSON C.R., SALAS J.A.;
RT "A cytochrome P450-like gene possibly involved in oleandomycin
RT biosynthesis by Streptomyces antibioticus.";
RL FEMS Microbiol. Lett. 127:117-120(1995).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; L37200; AAA9253.1; -.
DR HSSP; 000441; 10XA.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
DR PFAM; PF00067; P450; 1.
DR PRINTS; PR00359; BP450.
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.
FT BINDING 356 356 HEME (BY SIMILARITY).
SQ SEQUENCE 407 AA; 44957 MW; 04111C60 CRC32;
```

```
Query Match 16.1%; Score 350.5; DB 2; Length 407;
Best Local Similarity 28.4%; Pred. No. 8.9e-21;
Matches 105; Conservative 63; Mismatches 167; Indels 35; Gaps 11;
```

```
QY 60 GCHHATRGQLIRAYEDYRHFSSEC---PPIREAGEADFIPT-----SMDEPEQ 108
DB 50 GTAWLVTRMSDARIVLDSR-FSTAATDPYTPR-----MFTPEPBGVLAQDEPDH 101
QY 109 RQFALANQVGMPEVYKLENRIOELACSLIESLRPOG-CNFTEDYAEPPFIRFMILA 167
DB 102 TRLRRLVGKAFYARVEEMRPVRSIVSDLDQVANGSPADVLEFLAVPEPPAVICEL 161
QY 168 GLPEEDIPHLKYLTDO---TRPDGSMTEPAKALYDYL---PIEQRORQGTDAIS 221
DB 162 GVPLEDDRLRTESDAMLSSTR---LTAAEIQHVQODFMVYMGILVAORRDAPTEDLG 217
QY 222 IVANGVNGRPITSDAKRMCGILLVGCIDTVVNFSLFSMEFLAKSPENHOELIERPELI 281
DB 218 ALALATNDHLLTKGETLVNMGVSLLAGHETVNOITNLVHLLTERRKRESYLADPALV 277
QY 282 PAACEELLRRFSIYADG---RILTSDEYFPGVOLKKGQDILLPQMLSGLDERKNACPMHV 338
DB 278 PAAVEEMLRTPPLVSAGSPFRVATEDELSTVYVRAGEPCPVNHASANRDEEYVDHDEL 337
QY 339 DFRSQKVSHTTFGSHLCLGQHLARREIIVTLKWLTRIPDSIA-PGAQIOHKSG-IV 396
DB 338 DFHRRRNPHIAFGHAGHCIGAGQGLRLEQALVSALVRRPPTLDLAPVAGLKKWQGLMI 397
QY 397 SCVOALPLVW 406
DB 398 RGLERQIVSW 407
```

RESULT 7

Q59723 PRELIMINARY; PRT; 406 AA.

AC Q59723:

DT 01-NOV-1996 (Tremblrel. 01, Created)

DT 01-NOV-1996 (Tremblrel. 01, last sequence update)

DE CYTOCHROME P450 LIN (EC 1.14.14.1).

GN LINC.

OS Pseudomonas IncoGNita.

```
OC Bacteria; Proteobacteria.
RN [1]
RP SEQUENCE FROM N.A.
RA ROP J.D., GUNSAIUS I.C., SLIGAR S.G.;
RL Submitted (JAN-1994) to the EMBL/GenBank/DDJ databases.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; L23310; AAA25810.1; -.
DR HSSP; P33006; 1CPT.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
DR PFAM; PF00067; P450; 1.
DR PRINTS; PR00359; BP450.
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.
FT BINDING 355 355 HEME (BY SIMILARITY).
SQ SEQUENCE 406 AA; 45637 MW; EFAB4E0D CRC32;
```

```
Query Match 15.2%; Score 332; DB 2; Length 406;
Best Local Similarity 26.6%; Pred. No. 2.9e-19;
Matches 106; Conservative 73; Mismatches 176; Indels 44; Gaps 10;
```

```
QY 29 YNPSNLISAGVOEAMAVIOESNVDPDLVWTRCNGHWTATRGOLREAYEDYRH-FSSRCPF 87
DB 32 WNPESDSCGF---WAVLRKHDIIE-----VSROPLTSSAYENGCHRPENENEV 77
QY 88 IPRAGEAYDFIP-TSMDPPEORQFALANQVGMPEVYKLENRIOELACSLIESLRPOG 146
DB 78 GLTNAGSAAGVPEFISLDPPVHTQYRKVIMPALSPARLGDIEGRIRVRALEIRPLGE 137
QY 147 QCNFTEDYAEPPFIRIFMLAGL-PE-----EDIPHLKYLTDO---TRPDGS- 190
DB 138 EVDLVPLLSAPPLLTLLAEILGDPDCWELYNMTNAFVEDDEDEFKSPEDAKVIGEF 197
QY 191 MTEPAEAKALYDYLPIITIEQRORQGTDAISIYANGVNGVRPTTSDPAKRMCGILLVGL 250
DB 198 MGCCQ-----ELFESRRANPGPDITFLANAEINQPALRDFINLTTLVGCN 247
QY 251 DTVVNFSLFSMEFLANSPEHROELIERPELIPACEELLRRFSIYAD-GRILTSDEYFHG 309
DB 248 ETRRNISHTITVLSQOPDQMDILRQRPPELLKTATAEMVHNASPVLMRTAMEDPEIGG 307
QY 310 VOLKKGQILLPQMLSGLDERKNACPMHVDFPSQKVSHTTFGSHLCLGHLARREIIV 369
DB 308 QALAKGDKVLYLVASGNRDSVESDADRFPVTRTGVGHVFGSGOHVCVGSRLAEQRLV 367
QY 370 TLKEMLTRIPDFTIAPGAQIOHKSIGVQOALPLVWDP 408
DB 368 VFELLSTRVAKRFELCKSR-RFRSNFLNGLKNLVYLP 405
```

RESULT 8

Q92H01 PRELIMINARY; PRT; 420 AA.

AC Q92H01:

DT 01-MAY-1999 (Tremblrel. 10, Created)

DT 01-MAY-1999 (Tremblrel. 10, last sequence update)

DE CYTOCHROME P450.

GN TYLH1.

OS Streptomyces fradiae.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

RN [1]

RP SEQUENCE FROM N.A.

RA FOUCHES R., MELLADO E., DIEZ B., BARREDO J.L.;

RT "The left edge of the tylosin gene cluster from Streptomyces

RT fradiae.";

RL Microbiology 0:0-0(1999).

DR EMBL; AF055922; AAD12167.1; -.

DR HSSP; 000441; 10XA.

SQ SEQUENCE 420 AA; 45531 MW; 3A3474F7 CRC32;

Query Match 15.2%; Score 331.5; DB 2; Length 420;



```

Db      46  WLLTRHADVRLALADPGVSSHPGAPQPMRLNLAEM-----RAEHYLPGLLEMDPP  97
Qy      107  EORQFRLANOVCGMPYVDKLENNIOELACSLIESLRPO-GOCNFTEDAVEPPRIEML  165
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      98  DHTKRRLLTWMFTRRAIRKLEPRIEQVITETLDMAGOGSTYDLYOSFALPRLVICE  157
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      166  LAGLPEEDIPIHLKYLTDMQTRPDGSMTPFAEAKA-LYDYLPIIEORRQPGTDAISYA  224
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      158  LMGIYEEEREEFEMDVLTALQALDQTPETIGALGARMMNPMWKLAAKRAMPGDOLLSHA  217
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      225  NGQVNGRITSDAEAKRMCGLLVYGGDLTVVNFLEFSMEFLAKSPENHOELIERELIPAA  284
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      218  HDDPADPALTDLELAGIGVLMLAGHETSAMMLGVGTYTLLNNDQWALLRDISLIDRA  277
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      285  CEELLRRRESLVAOD--RIITSDVEFHGVOIAKKGOITLLPWLSCGLDERKNACPMHVDFSR  342
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      278  VEEILRHQTIYVOGGLPRQVYTDMDIELAGHQVYTGSSLASLPAAARNDPAVFPDPDRDLITR  337
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      343  QKVSHTTTGHGSHLCLGQHLARREIYTLKEMLTRIDPFSIAPGAQ-IQHSKG-IYSGVO  400
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      338  EHNPHLAGHGCHICLQELARVEMRQAMRGVLVTRPFGELRMAAAPEDIIRMRDQIYGVY  397
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      401  ALPLVMDPA 409
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      398  NLPTVMDPA 406
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT	11			
032460				
ID	032460	PRELIMINARY;	PRT;	411 AA.
AC	032460;			
DT	01-JAN-1998 (TREMBLrel. 05, Created)			
DT	01-JAN-1998 (TREMBLrel. 05, Last sequence update)			
DT	01-NOV-1999 (TREMBLrel. 12, Last annotation update)			
DE	ORF 10.			
OS	Actinomadura hibisca.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
OC	Actinomycetales; Streptosporangineae; Thermomonosporaceae;			
OC	Actinomadura.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=PI57-2;			
RX	MEDLINE; 97480928.			
RA	DAIJI T., HAMANO Y., IGARASHI Y., FURUMAI T., OKI T.;			
RT	"Cloning and nucleotide sequence of the putative polyketide synthase			
RT	genes for pradiacin biosynthesis from Actinomadura hibisca";			
RL	Biosci. Biotechnol. Biochem. 61:1445-1453(1997).			
DR	EMBL; D87924; BAA23153.1, -.			
DR	HSSP; O00441; 10XA.			
DR	PFAM; PF00067; p450, 1.			
DR	PRINTS; PR00359; BP450.			
SO	SEQUENCE	411 AA;	44860 MW;	98A628B6 CRC32;

Query Match	15.08;	Score 327;	DB 2;	Length 411;
Best Local Similarity	26.48;	Pred. No. 7.5e-19;		
Matches 114; Conservative	61;	Mismatches 199;	Indels 58;	Gaps 16;

[illegible]

RESULT	12			
09X9P7				
1D	09X9P7	PRELIMINARY;	PRT;	410 AA.
AC	09X9P7;			
DT	01-NOV-1999	(TREMBLrel. 12, Created)		
DT	01-NOV-1999	(TREMBLrel. 12, Last sequence update)		
DT	01-NOV-1999	(TREMBLrel. 12, Last annotation update)		
DE	N1F7 PROTEIN.			
GN	N1F7.			
OS	Streptomyces tendae.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
OC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-TUE901;			
RA	BRUNTER C., LAUER B., SCHWAB W., MOHRLE V., BORMANN C.;			
RT	"Molecular characterization of co-transcribed genes from Streptomyces			
RT	tendae TUE901 involved in the biosynthesis of the peptidyl moiety of			
RL	the peptidyl nucleoside antibiotic nikkomycin."			
RL	Mol. Gen. Genet. 0:0-0(0).			
SO	EMBL; Y18574; CAB46536.1; -			
SO	SEQUENCE 410 AA; 45884 MW; 9B52CB74 CRC32;			

Query Match	14.8%	Score 322:	DB 2:	Length 410:
Best Local Similarly	27.3%	Pred. NO. 1.9e-18:		
Match 99, Conservative	64:	Mismatches 179:	Indels 20:	Gaps 9:

```

QY 63 W1ATFGOLIREAYDYRHFSESC -PFTP-----REAGAYDFIPN-----SMDPEOR 109
Db 47 WLVLKHDLARLKLADPRVSADRLPAFPFGLTAEQFATERYVRRLSTRSMHLDGDEHG 106
QY 110 QFRALANQVGMPPVDKLENNIOELA -CSLIESRPGQCNCFEDVAEPPIPIFMILNG 168
Db 107 AHKILGESSLRKIALRPPVQCTIVDRSIDEMALAPQADLVEHVSQAVPSLVICELLG 166
QY 169 LPEDDIPHL -KYLLDOMTRPDGSMTFEAKAEALDYDILPII -EQRORPGCTDAISIVANG 226
Db 167 VPHQRDRDFHEMAGMLVSRSYSINERAAASDALNDLPLEDLVTEKEREPDIDLIGRLAR 226
QY 227 QVNGRPITSDCAKMCGLLYVGLDTPVNFSLTSSMEFLAKSPENROBLERPELIPACE 286
Db 227 NRRTPVMTHEIVGTAVMLLIAGHOTANNISLGVALLENPEHKARIADPSLLPDATE 286
QY 287 ELLRREFSLV -ADRRILITSDYEFNGVQLKKGDOILLPQMLSGLDERKNACPMHVDSSROK 344
Db 287 EMLTRFSVENAPARVATDEDEIGGVITRDEGLVSGLAADWDDEVEFHDRDLDERGA 346
QY 345 VSHTFEGSHILCGOHLARREIIVTLKEMLTRIPDSIA -PGAOJOHKSGI -VSGVOAL 402
Db 347 RHHAFGCVGHCQGLQMLARVELEYETLLRRVPGSLAVPAEELPYKDDAGIYGIYRY 406
QY 403 PL 404
Db 407 PV 408

```

RESULT 13  
059831 PRELIMINARY; PRT; 410 AA.  
AC 059831;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
DE CYTOCHROME P450 SCA-2 (EC 1.14.14.1).  
OS Streptomyces carboxylus.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomycetes.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SANK 62585;  
RX MEDLINE; 96001248.  
RA WATANABE I., NARA F., SERIZAWA N.;  
RT "Cloning, characterization and expression of the gene encoding  
cytochrome P-450sca-2 from Streptomyces carboxylus involved in  
production of pravastatin, a specific HMG-CoA reductase inhibitor.";  
RL Gene 163:81-85(1995)  
DR EMBL; D30815; BAA06492.1; -.  
DR HSSP; Q00441; 10XA.  
DR PFAM; PF00067; P450; 1.  
DR PRINTS; PR00359; BP450.  
SQ SEQUENCE 410 AA; 45049 MW; 2DF70C8E CRC32;

Query Match 14.7%; Score 319.5; DB 2; Length 410;  
Best Local Similarity 25.9%; Pred. No. 3e-18;  
Matches 93; Conservative 63; Mismatches 184; Indels 19; Gaps 7;

QY 63 WIATRGQLIREAYEDYR-----HFSSECPRIIP--REAGEADFTPTSDPPEQROFR 112  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
DB 56 WVTYKHEAARLLADPRLSSDRHADPAPSPRKAERQSGPAF---IGMDPEHGSTR 111  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
QY 113 ALANQVGMPEVDKLENIROELACSLIESLRPOGQCNFEDYAEPPRIEMLAGLEPE 171  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
DB 112 RMTISEFTVARIKGMRPDVERIVNGFIDMLACPTADIVSQFALPSPWVTCIMGLVPR 171  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
QY 172 EDIPHLKYLTDQMTRPDGSMTFAEAKALYDYLPIIEQRQKPGTDAISIVANGQNGR 231  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
DB 172 ADHEFFQDASKRLVQAVDADSAVARDDFERVYDGLITKLESEPGTGLKLVHQLADG 231  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
QY 232 PIYSDEAKRMCGLLVVGGLDTVAVNFIQSMFELAKSPENROELIERPELLPAACEELLRR 291  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
DB 232 EIDRAELISTALLLVAGHETTSMTSLSVTTLEHPDOHAALRADPSLVPGAAVEELLRV 291  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
QY 292 FSL--VADGRILTSDEYFEGHVGOLKGDQILLPQMLSGLDERKNACPMHVDPSRQKVSHTT 349  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
DB 292 LAIDINGGRATADITDGLIRAGCGVITYTIANRDSVFPNPDLVDHRSARHHLIS 351  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
QY 350 FGHSHLCLGOLARREIITVLEKMLTRIPDSIA-PGAQIQKSG-IVSGVALPLVW 406  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
DB 352 FGYGVHCGICGNLARLELEVILVTLFDRIPRLAVPVEQTLTGTTIGVNEPLPTW 410  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 14  
09XDB0 PRELIMINARY; PRT; 400 AA.  
ID 09XDB0;  
AC 09XDB0;  
DT 01-NOV-1999 (Tremblrel. 12, Created)  
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
DE CYTOCHROME P450.  
GN PIPA.  
OS Mycobacterium smegmatis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MC2155;

RX MEDLINE; 99287823.  
RA POUJIN P., DUCROCQ V., HALLIER-SOULIER S., TRUFFAUT N.;  
RT "Cloning and characterization of the genes encoding a cytochrome P450  
(Pip) involved in piperidine and pyrrolidine utilization and its  
regulatory protein (P4PR) in Mycobacterium smegmatis mc2155.";  
RL J. Bacteriol. 181:3419-3426(1999).  
DR EMBL; AF102510; MAD28344.1; -.  
SQ SEQUENCE 400 AA; 44747 MW; C10DD01A CRC32;

Query Match 14.6%; Score 318; DB 2; Length 400;  
Best Local Similarity 27.0%; Pred. No. 3.9e-18;  
Matches 116; Conservative 57; Mismatches 170; Indels 86; Gaps 14;

QY 26 FDMYNP---SNLSAGVOEA---W-----AVLOESVNPDLV-----WTGNG 60  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
DB 12 FDTVDPAFSTSDSEVHEKRSWATTPYGLAVRLYRDVNRILNPKLRGSSAMPANH 71  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
QY 61 -----GHWITATRGQLIREAYEDYRHFSSECPRIIPREAGEAYDPTSDPPEQROFRAL 114  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
DB 72 VTGEPFAEWFAF-----WLNKKE-----EHHRLRL 99  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
QY 115 ANOVGMPEVDKLENIROELACSLIESLRPOGQCNFEDYAEPPRIEMLAGLEPE 174  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
DB 100 MNPAPSPKLGSLVPRQALNELIDNFAEPDCEFEVSEFAEYARVIAIMGLPEBE- 158  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
QY 175 PHUKYLTQ-----MTRPDGSMTFAEAKALYDYLPIIEQRQKPGTDAISIVAN 225  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
DB 159 --WVISTESATIGLAVTLREDLPKIEAVQRLYEXSDLEADRANPDDPMTLVN 216  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
QY 226 GQVNGRP-----ITSDEAKRMCGLLVVGGLDTVAVNFIQSMFELAKSPENROELIERPELL 281  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
DB 217 A---SRPDDGRLSKELRDALLILFGFDTTRNQLGLAMTFKKNHDPKMLLBERDGL 273  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
QY 282 PACEELLRRPSLYA-DRIITSDYEFHVGOLKGDQILLPQMLSGLDERKNACPMHVD 340  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
DB 274 GKAVEEYMRVPTVRYWYREVVEDEFEVGYTLKAGTYVHLTSESAGDPR--VEPGEDI 331  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
QY 341 SROKVSHTTEGHSHLCLGOLARREIITVLEKMLTRIPDSIAPGAQIQKSGIVSQ 400  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
DB 332 TAERKPHFEGGVHCHGHPVARSMSALPRLARRLDPEHLPGATWLPDSC-NTGPN 390  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
QY 401 ALPLVWDP 409  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
DB 391 TLRPGFTPA 399  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 15  
032927 PRELIMINARY; PRT; 310 AA.  
ID 032927;  
AC 032927;  
DT 01-JAN-1998 (Tremblrel. 05, Created)  
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)  
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
DE CYTOCHROME P450 (FRAGMENT).  
GN MCB1788.51C.  
OS Mycobacterium leprae.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA SKELTON J., CHURCHER C.M.;  
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;  
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA MEDLINE; 93188700.  
RX EIGLMETTER K., HONORE N., WOODS S.A., CAUDRON B., COLE S.T.;  
RT "Use of an ordered cosmid library to deduce the genomic organization  
of Mycobacterium leprae.";



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 12:35:16 ; Search time 76.18 Seconds

(Without alignments)  
128.722 Million cell updates/sec

Title: US-09-246-451-13

Perfect score: 2177

Sequence: 1 TTERIQSNANLAPLPHPVE.....IVSGVQALPLWDPATRAV 414

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2122	97.5	405	1	Y04128	Pseudomonas putida
2	1185	54.4	446	1	Y04126	Bacterial and mamm
3	362	16.6	411	1	R60777	Mycelamycin IV hyd
4	337	15.5	411	1	W54389	Actinomadura hibis
5	328.5	15.1	587	1	W33374	S. fradiae tylosin
6	328	15.1	406	1	R11349	Cytochrome enzyme
7	324.5	14.9	398	1	W11585	Streptomycetes prist
8	322	14.8	403	1	R11350	Cytochrome enzyme
9	321	14.7	408	1	R77867	S. clavuligerus OR
10	314	14.4	412	1	R38309	Sequence of the P4
11	311.5	14.3	410	1	R51368	Protein containing
12	294.5	13.5	404	1	R14724	6-hydroxylase enco
13	190	8.7	398	1	R47521	Vitamin D hydroxyl
14	161	7.4	422	1	W36128	Daunomycin C-14 hy
15	161	7.4	474	1	W36132	N-terminal modifie
16	156	7.2	422	1	W00729	Daunorubicin 14-hy
17	152	7.0	533	1	R15057	Cytochrome P450C25
18	142	6.5	494	1	R62825	Human steroid-21-h
19	135.5	6.2	1169	1	R76544	Mitochondrial cyto
20	133	6.1	518	1	W67616	A. nidulans phenyl
21	131	6.0	512	1	W93216	Human cytochrome p
22	129	5.9	508	1	W35711	Chrysanthemum flavo
23	129	5.9	512	1	R72365	Human auxillary cy
24	128	5.9	512	1	R93172	Human cytochrome p
25	128	5.9	512	1	R72366	Human auxillary cy
26	128	5.9	512	1	W00652	Cytochrome P4501A1
27	128	5.9	512	1	R93173	Human cytochrome p
28	125	5.7	524	1	P70577	Rat hepato-cytochr
29	123	5.6	503	1	R72363	Human cytochrome p
30	123	5.6	503	1	R81464	Human derived cyto
31	123	5.6	503	1	R93170	Human cytochrome p
32	123	5.6	503	1	Y05202	Human CYP3A4 prote
33	123	5.6	512	1	R72364	Human auxillary cy

34	123	5.6	512	1	R93171	Human cytochrome p
35	122.5	5.6	523	1	R59291	Rat liver cytochro
36	122.5	5.6	898	1	P61030	Entire coded sequ
37	122.5	5.6	898	1	P61082	Complete translati
38	122.5	5.6	899	1	P61056	translation of pla
39	122.5	5.6	1144	1	P81334	Expression prod. o
40	122.5	5.6	1150	1	P81335	Expression prod. o
41	122.5	5.6	1150	1	P81337	Expression prod. o
42	122.5	5.6	1162	1	P81336	Expression prod. o
43	121	5.6	514	1	W34539	Cytochrome P450 cc
44	120.5	5.5	493	1	R72362	Human cytochrome p
45	120.5	5.5	493	1	R81467	Human derived cyto

ALIGNMENTS

last-processed

RESULT	1
ID	Y04128
AC	Y04128; Y04128 standard; Protein; 405 AA.
DE	11-JUN-1999 (first entry)
KW	Pseudomonas putida cytochrome P450 protein P450-cam.
KM	Bacterial; mammalian; cytochrome P450; chimeric; fusion protein;
KW	oxidise; hydrocarbon; carbon-hydrogen bond; hydroxylating;
OS	Pseudomonas putida; environmental pollutant.
PN	MO9908812-AI.
PD	25-FEB-1999.
PF	17-AUG-1998; U16979.
PA	20-AUG-1997; US-056754.
PI	(UYRP ) UNIV ROCHESTER.
DR	WPI: 99-190131/16.
DR	N-PSDB: X19926.
PT	New P450 fusion proteins - comprising a portion of a bacterial
PT	cytochrome P450 protein and a portion of a mammalian cytochrome P450
PT	protein
PS	Disclosure: Page 12-13: 51pp; English.
CC	The present invention describes a fusion proteins comprising a portion
CC	of a bacterial cytochrome P450 protein and also a portion of a mammalian
CC	cytochrome P450 protein. The fusion protein can oxidise hydrocarbons or
CC	any compound having a carbon-hydrogen bond. The fusion protein can be
CC	used for hydroxylating a compound to be oxidised. It can also be used in
CC	the bioremediation of an environmental pollutant. Since the fusion
CC	protein is soluble, it can be subject to structural elucidation by X-ray
CC	crystallography for designing functional proteins. It can be readily
CC	expressed in soil bacteria to facilitate bioremediation. The present
CC	sequence represents Pseudomonas putida cytochrome P450 protein P450-cam
CC	from the present invention.
SQ	Sequence 405 AA:
Query Match	97.5%; Score 2122; DB 1; Length 405;
Best Local Similarity	99.5%; Pred. No. 2e-210;
Matches 403; Conservative	1; Mismatches 1; Indels 0; Gaps 0;
Db	10 NLAPLPHPVEHLVDFDPMTPSNLSAGVOBAMVLDGSNPDLVWTRCNGHMIATRCQ 69
Db	1 NLAPLPHPVEHLVDFDPMTPSNLSAGVOBAMVLDGSNPDLVWTRCNGHMIATRCQ 60
QY	70 LIRAYEDYRHFSSECPPIREAGEAYDFPTSMDDPEOROFRLANOVGMPVVDKLN 129
Db	61 LIRAYEDYRHFSSECPPIREAGEAYDFPTSMDDPEOROFRLANOVGMPVVDKLN 120
QY	130 RIQELASLIESLRPOGCFNTEDYAEFPPIRIMLAGLPEEDIPHLKYLTDDMTTPDG 189
Db	121 RIQELASLIESLRPOGCFNTEDYAEFPPIRIMLAGLPEEDIPHLKYLTDDMTTPDG 180
QY	190 SMTPAEAKKALYDLIPITDOROKPGTDAISIYANGVNGRPITSDPAKRMGCLLVGG 249
Db	181 SMTPAEAKKALYDLIPITDOROKPGTDAISIYANGVNGRPITSDPAKRMGCLLVGG 240

QY 250 LDVYVNFSLFSMEFLAKSPHEHOELIERPERIPACCELLRRFSLVADGRILTSDEYEFHG 309  
 |||||  
 Db 241 LDVYVNFSLFSMEFLAKSPHEHOELIERPERIPACCELLRRFSLVADGRILTSDEYEFHG 300  
 |||||  
 QY 310 VOLKKGQIILLPQMLSGDLDERKNACPMHVDSPROKVSHTTGFHSGSHLCLGOHLARRETIY 369  
 |||||  
 Db 301 VOLKKGQIILLPQMLSGDLDERKNACPMHVDSPROKVSHTTGFHSGSHLCLGOHLARRETIY 360  
 |||||  
 QY 370 TLKEMLTRIPDFSIAPGAQIOHKSGIVSGVQALPLVMDPATTKAV 414  
 |||||  
 Db 361 TLKEMLTRIPDFSIAPGAQIOHKSGIVSGVQALPLVMDPATTKAV 405  
 |||||

RESULT 2  
 ID Y04126  
 AC Y04126: standard; Protein; 446 AA.  
 DT 11-JUN-1999 (first entry)  
 DE Bacterial and mammalian chimeric cytochrome P450 protein.  
 KW Bacterial; mammalian; cytochrome P450; chimeric; fusion protein;  
 KM oxidase; hydrocarbon; carbon-hydrogen bond; hydroxylating;  
 OS bioremediation; environmental pollutant.  
 OS Synthetic.  
 PN WO908812-A1.  
 PD 25-FEB-1999.  
 PF 17-AUG-1998; 016979.  
 PR 20-AUG-1997; US-056754.  
 PA (UYRP ) UNIV ROCHESTER.  
 PI Jones JP, Shimoji M;  
 DR WPI: 99-190131/16.  
 DR N-PSDB: X19916.  
 PT New P450 fusion proteins - comprising a portion of a bacterial  
 cytochrome P450 protein and a portion of a mammalian cytochrome P450  
 protein.  
 PS Claim 24: Page 6-8; 51pp; English.  
 CC The present sequence is a fusion proteins comprising a portion of a  
 bacterial cytochrome P450 protein and also a portion of a mammalian  
 cytochrome P450 protein. The fusion protein can oxidise hydrocarbons or  
 any compound having a carbon-hydrogen bond. The fusion protein can be  
 used for hydroxylating a compound to be oxidised. It can also be used in  
 the bioremediation of an environmental pollutant. Since the fusion  
 protein is soluble, it can be subject to structural elucidation by X-ray  
 crystallography for designing functional proteins. It can be readily  
 expressed in soil bacteria to facilitate bioremediation.  
 SQ Sequence 446 AA;

Query Match 54.4%; Score 1105; DB 1; Length 446;  
 Best Local Similarity 58.0%; Pred. No. 9e-114;  
 Matches 237; Conservative 32; Mismatches 92; Indels 62; Gaps 11;  
 QY 10 NLAPLRPHVEHILVDFDMYNPNSLSAGVOEAMAVLOESNVPLDVMTRCGHMIATRGQ 69  
 |||||  
 Db 1 NLAPLRPHVEHILVDFDMYNPNSLSAGVOEAMAVLOESNVPLDVMTRCGHMIATRGQ 60  
 |||||  
 QY 70 LIREAVEDYRHFSSECFPIREAGEAYDFTPTSMDPREOROFALANOVVMPPVDKLEN 129  
 |||||  
 Db 61 LIREAVEDYRHFSSECFPIREAGEAYDFTPTSMDPREOROFALANOVVMPPVDKLEN 120  
 |||||  
 QY 130 RIQELACSLTESLRPOQCNCFTEDYAEPPRIIRIFMLAGIPREEDIPHLKLTIDQMTRPDG 189  
 |||||  
 Db 121 RIQELACSLTESLRPOQCNCFTEDYAEPPRIIRIFMLAGIPREEDIPHLKLTIDQMTRPDG 180  
 |||||  
 QY 190 SMFPAEKALYDYLPIIIOOROKPCT-----DAISIYANGOVNRP-----ITSDE-- 237  
 |||||  
 Db 181 SMFPAEKALYDYLPIIIOOROKPCT-----DAISIYANGOVNRP-----ITSDE-- 240  
 |||||  
 QY 238 AKMFGILLVGLDVTYVNFSLFSMEFLAKSPHEHOELIERPERI----- 281  
 |||||  
 Db 241 AVDLFG----AGTETSTLTALVALLLKLKPEVTAKVQEEIERYGNRSPCQDRSHMP 296  
 |||||  
 QY 282 --PAACEELLRRSYVADG--RLTSDYEFHGVOLKKGDQIIL-----POM 323  
 |||||

Db 297 YTDVAVHEVORYIDLPTSLPHAVTCDIKFRNYLIPKGTILISTSLVLDHNKEFPNDEM 356  
 |||||  
 QY 324 LSG---LDERKNACPMHVDSPROKVSHTTGFHSGSHLCLGOHLARRETIYVLKEMLTRIPD 380  
 |||||  
 Db 357 FDRHNHFLDEGNN-----FKSKTY-FMPSAGKRIVGEGALMGELFLFULTSLONFNL 408  
 |||||  
 QY 381 FSIAPGAQIOHKSGIVSGVQALP 403  
 |||||  
 Db 409 KSLVDPRKNDL-TTPVYNGFASVP 430  
 |||||

RESULT 3  
 ID R60777  
 AC R60777: standard; Protein; 396 AA.  
 DT 21-JUN-1995 (first entry)  
 DE Mycinamicin IV hydroxylating protein.  
 KW Mycinamicin; hydroxylase; macrolide; antibiotic; Micromonospora;  
 KM Micromonospora griseorubida.  
 OS Micromonospora griseorubida AT11725CN3.  
 PN J06253853-A.  
 PD 13-SEP-1994.  
 PF 09-MAR-1993; 047638.  
 PR 09-MAR-1993; JP-047638.  
 PA (ASAH ) ASAH KASEI KOGYO KK.  
 DR WPI: 94-328997/41.  
 DR N-PSDB: 073674.  
 PT DNA encoding a protein having mycinamicin IV hydroxylating  
 activity - for prodn. of mycinamicin, a macrolide antibiotic  
 PS Claim 1; Page 12-14; 23pp; Japanese.  
 CC The amino acid sequence of a protein having mycinamicin IV hydroxylase  
 (MH) activity. The gene encodes a protein of 396 a.a. The DNA was  
 obtained from the macrolide antibiotic-producing bacterium Micromonospora  
 griseorubida A11725CN3/PHY507. The gene was isolated from the plasmid  
 PHY507. The protein encoded by this plasmid can be used to produce  
 mycinamicin IV in PHY507-deficient Micromonospora strains.  
 SQ Sequence 396 AA;

Query Match 16.6%; Score 362; DB 1; Length 396;  
 Best Local Similarity 28.5%; Pred. No. 5.2e-29;  
 Matches 105; Conservative 57; Mismatches 168; Indels 38; Gaps 8;  
 QY 63 WIATRGQILIREAVEDYRHFSSECFPI-----PREAGEAYDFTPTSMDPREOROFRA 113  
 |||||  
 Db 43 WLVTYR-----YEDYRAVLGDSGFVGRPSMTDERPREVWKGLSMDPREHSLRR 95  
 |||||  
 QY 114 LANOVGMPPVDKLENRIQELACSLTESLRPOQ-CNFTEDYAEPPRIIRIFMLAGIPRE 172  
 |||||  
 Db 96 LVVKAFTARRAESLRPRAREIAHELYDQMATGCPADLVAMFAROLPVRVICELGVPSA 155  
 |||||  
 QY 173 DIHKLKLTIDQMTRPDQSM-----TFEAKKALYDYLPIIIOOROKPCTDAISIV 223  
 |||||  
 Db 156 D-----HDFTRWSCAGFSTAETVAEEQEAQAVAYMGDLIDRRKEPTDDLVSA 208  
 |||||  
 QY 224 ANQOVNRPITSDAKRMFGLLVGLDVTYVNFSLFSMEFLAKSPHEHOELIERPERIPA 283  
 |||||  
 Db 209 VQARDQDSSLSEBELDLDAIGLLVAGYESYTTQIADRVYLLMTRERLROLDRLPELIPS 268  
 |||||  
 QY 284 ACEELLRRSL---VADGRILTSDEYEFHGVOLKKGDQIILPQMLSGDLDERKNACPMHVD 340  
 |||||  
 Db 269 AVEELTRWPLDVGTAFFRVAVEDVTLRGVTIRAGEVYLASTGAANDQAOFPDADRIDV 328  
 |||||  
 QY 341 SRQKVSHTTGFHSGSHLCLGOHLARRETIYVLKEMLTRIPDFSTA--PQAOQHKSG--IVSG 398  
 |||||  
 Db 329 DRTPRNHLGFGHGVNHCGLARVELQVALLQRLPGLIGIPETQIRWSEGMILRG 388  
 |||||  
 QY 399 VQALPLTW 406  
 |||||  
 Db 389 PLELPVTW 396  
 |||||  
 RESULT 4





DR N-PSDB; Q11126.  
PT DNA encoding cytochrome P450 enzymes - and electron donating  
PT iron-sulphur proteins, used to confer herbicide resistance to  
PT plants and microorganisms  
PS Claim 13: page 151: 224pp: English.  
CC This cytochrome P450 enzyme, P450SU1 is expressed alongside the iron  
CC sulphur protein P450-B, by a DNA sequence contained in a recombinant  
CC plasmid. Host Streptomyces species are transformed with the plasmid  
CC and are used to coat a plant seed to transform the plant. The res-  
CC ultant transformants are resistant to herbicides.  
CC See also Q11127.  
SQ Sequence 406 AA:

Query Match 15.1%; Score 328; DB 1; Length 406;  
Best Local Similarity 26.3%; Pred. No. 1.7e-25;  
Matches 104; Conservative 74; Mismatches 193; Indels 24; Gaps 11;

QY 31 PSNLSAGVQ--EAMAVLQESNVPDLVWTRCGNH-WIATRGQLREAYEDYR----- 79  
||| | :  
DB 17 PSNRSCPYYQLPDGYAQLRDPGLHRYTLXDGQAWVVTKEHAAARKLLGDPRLSSNRTDD 76  
||| | :  
QY 80 HESSECPFR--REAGRAYVFIPTSMNDREORQFRALANQVGVVYDKLENRIQELACS 137  
:  
DB 77 NFRATSPREAVRESPOAF---IGDPRRHGTRRRRTISEFTVVKIKGMRPEVEVHVG 132  
:  
QY 138 LIESLRPOG-OCNFTEDYAEFPPIRIFMLAGLREEDIPHLKYLTQDMTRPDSMTFAEA 196  
:  
DB 133 FLDEMLAGTADLVSGFALVPVSMVICRLGVYADHEFFQDASKRLVSTDAQSALTA 192  
:  
QY 197 KEALYDYLPIIQRORRGTDAI-STVANGQVNGRPISTDEAKRMFGLLVGGDLTVN 255  
:  
DB 193 RNLGAGLDLITQOFQEPGAGLVGALVADOLANGE-IDREELISTAMLLIGHETTAS 251  
:  
QY 256 FLSEMFELAKSPENHQLERPERIPRACEELIRFSL--VADGRITSDYEHNGVQLK 313  
:  
DB 252 MTLSTVLTLDHDEQYALADRSILVGAAYELRYLATIDINGRAVATADIEVGHILR 311  
:  
QY 314 KGDQILLPOMLSGLDERKNACPMHNVDFSROKVSHTTGHGSHLCLGHLARREIIVTLKE 373  
:  
DB 312 AGSGVIVVNSIANRDGVYEDRDALDINRSARHNLIAAGFVHQCLOGNLARLEVEYLNA 371  
:  
QY 374 WLTRIPDESLA-PGAQIQHKSG-IVSGVALPLVW 406  
:  
DB 372 LMDRVPTLRILAVPEQVLVLRPTTIGGVNELPYTW 406  
:  
RESULT 7

W11585  
ID W11585 standard; Protein; 398 AA.  
AC W11585;  
DT 02-APR-1997 (first entry)  
DE Streptomyces pristinaespiralis snbr gene product.  
KW Streptogramin B; antibiotic; biosynthesis; pristinaamycin;  
KW virginiamycin; pibecolic acid; cyclodamination; papa; snba; snbf;  
KW piba; 3-hydroxypibecolic acid; hydroxylation.  
OS Streptomyces pristinaespiralis.  
PN MO9601901-A1.  
PD 25-JAN-1996.  
PF 04-JUL-1995; F00889.  
PR 08-JUL-1994; FR-008478.  
PA (RHON ) RHONE POULENC RORER SA.  
PI Barriere JC, Blanc V, Blanche F, Crouzet J;  
PI Debussche L, Paris JM, Thibaut D, Bamas-Jacques N;  
PI Dutuc-Rosset G, Famechon A;  
DR WPI: 96-097631/10.  
DR N-PSDB; T58555.  
PT New streptogramin B derivs. useful as antibiotics - produced by new  
PT mutants of Streptomyces having altered genes for streptogramin B  
PT biosynthesis  
PS Example 1: Page 113-114; 146pp: French.  
CC The papa gene of S.pristinaespiralis is involved in the biosynthesis  
CC of 4-dimethylamino-L-phenylalanine (DMPAPA), a precursor for

CC pristinaamycin IA. Upstream of the papa gene, on the complementary  
CC strand, is the snba gene coding for 3-hydroxypibecolic acid-AMP ligase.  
CC The region between these two genes was sequenced and two open reading  
CC frames were identified. The first (piba) decodes to an amino acid  
CC sequence with homology to ornithine cyclodaminase from Agrobacterium  
CC tumefaciens. The piba gene product is likely to catalyse the cyclo-  
CC deamination of lysine, leading to production of pibecolic acid.  
CC Mutations in the piba gene were shown to affect pibecolic acid  
CC synthesis but not the synthesis of 3-hydroxypibecolic acid. The second  
CC open reading frame (snbf) could be decoded to give a product with  
CC homology to hydroxylases of the cytochrome P450 type.  
CC Disruption of the piba and snbf genes can be used to produce  
CC strains of S.pristinaespiralis which are unable to produce the  
CC antibiotic pristinaamycin I but which may be able to produce new,  
CC modified forms of it.  
SQ Sequence 398 AA:

Query Match 14.9%; Score 324.5; DB 1; Length 398;  
Best Local Similarity 29.0%; Pred. No. 3.9e-25;  
Matches 106; Conservative 59; Mismatches 159; Indels 41; Gaps 12;

QY 74 AYEDYR-----FSSECPFRPAGRAYDPIPTSMNDREORQFRALANQVGM 121  
:  
DB 36 AFHVFRADVLTVASDPGVYSQSLRSPGQALSEQLSIDPPIHRLRRVLSQAFTR 95  
:  
QY 122 PVYDKLENRIQELACSLIESLRPOGOC-NFTEDYAEFPPIRIFMLAGLREEDIPHLKYL 180  
:  
DB 96 RTVADLEFRVTELAGQLLDV--DGFEDLVADFAPRPLVPIYAEILLGVPRADRTLFRSW 133  
:  
QY 181 TDQMT-----PDGSMTFARAKALYDYLPIIQRORRGTDAISIVA 224  
:  
DB 154 SDRMLQMVADPADMQFGDDADEDYQRLVKEPRAMHAYLDHVTDRBRANDLSALV 213  
:  
QY 225 NGQVNGRPISTDAKRMFGLLVGGDLTVNFI-SFSMEFLAKSPENHQLERPER--IP 282  
:  
DB 214 AARVEGERLDEQIVFEGALLMAAGHVSMLGNVYLCKDHP--RAEAAARADSLIP 271  
:  
QY 283 ACEBELLR-RFSLVADGRITSDYEFHGVOLKKGDOILLPOMLS-GLDERKNACPMHNVDF 340  
:  
DB 272 ALIEEVLRLRPPIVWARVTTKPTVLAGTTIRAG-RKVVYSLLSANHDEQVFPDRLHLD 330  
:  
QY 341 SRQKVSHTTGHGSHLCLGHLARREIIVTLKEMLRIPDPSIAPGAQIQ-HKSGIVSGV 399  
:  
DB 331 ARRG-RQIARHGHIHVCAPLARLEGRILALFDRFPDPSPTDGAKLRYHDGLF-GV 388  
:  
QY 400 QALPL 404  
:  
DB 389 KNLPL 393  
:  
RESULT 8

R11350  
ID R11350 standard; Protein; 403 AA.  
AC R11350;  
DT 05-JUN-1991 (first entry)  
DE Cytochrome enzyme P450SU2.  
KW Cytochrome P450; P450SU1; P450SU2; herbicide resistance.  
OS Streptomyces griseolus.  
PN WO9103561-A.  
PD 21-MAR-1991.  
PF 27-AUG-1990; U04785.  
PR 11-SEP-1989; US-405605.  
PR 12-JAN-1990; US-464489.  
PR 23-AUG-1990; US-569781.  
PA (DUPQ ) DU PONT DE NEMOURS CO.  
PI Dean C, Harder PA, Leto KJ, Lichteuer FT, Odell JT;  
PI O'Keefe DP, Omer CA, Romesser JA;  
DR WPI: 91-102077/14.  
DR N-PSDB; Q11127.  
PT DNA encoding cytochrome P450 enzymes - and electron donating  
PT iron-sulphur proteins, used to confer herbicide resistance to  
PT plants and microorganisms

PS Claim 15; page 158; 224pp; English.  
CC This cytochrome P450 enzyme, P450S02 is expressed alongside the iron  
CC sulphur protein Fes-A, by a DNA sequence contained in a recombinant  
CC plasmid. Host Streptomyces species are transformed with the plasmid  
CC and are used to coat a plant seed to transform the plant. The res-  
CC ultant transformants are resistant to herbicides.  
CC See also Q11126.  
SQ Sequence 403 AA;

Query Match 14.8%; Score 322; DB 1; Length 403;  
Best Local Similarity 28.4%; Pred. No. 7; le-25;  
Matches 103; Conservative 64; Mismatches 166; Indels 30; Gaps 13;

QY 63 WITRQGLIEAVEDEVHFSSEC-----PPIPRAGAYDFIPI--SMDPPEQKQFALA 115  
DB 52 WLVTRHODVRAVLGDPK-ESADAHRTGCPFLTMGGRITICTNPFLMDDPEHARLRML 110  
QY 116 NOVVGMPVDKLENRIQELACSLIESLRP-QGOCNFTEDYAEPPPIRIFMLAGLPEDI 174  
DB 111 TADFIYKKEVAMREVGRLADLDVDRMTGRTSADLVTERALPLSLVICLLGVYEDH 170  
QY 175 PHL-----KYLTDOMTPRDSGMTFAEAKALYDILPIIEOROKPGTDAIS-IVANGOVN 229  
DB 171 AFROERSRVLLTRSTPE---EVRADODELLEYLARLARKRERPDAAISRVLARGELD 227  
QY 230 GRPITSDIEMARFGLLVGLDVTNVNLFSSMEFLAKSPERHQLERPERIPACCELL 289  
DB 228 DTQIAT-----MGRLLVVARGEIDDTQIATMGLLRNPOLARLRAEPALVGVAVELL 282  
QY 290 RRESLVADG--RLITSDYEFHGVOLKKGQDILLPOMLSGLDERKNACP--MHVDFSRQV 345  
DB 283 RYLTIYVNGVPRIATEVVLGGRTIAGEGVLC--MISSANRAEVPFGDDLDVARDAR 340  
QY 346 SHPTFGSHLCIGQHLARREIIVTKEMLRIPDSIA-PGAQIQKSGI-VSGVALP 403  
DB 341 RHVAFGGVHQCIGQPLARVELQIAIETLLRLRLDLRLAVPHEIIPFGDMATGVHSLP 400  
QY 404 LVW 406  
DB 401 IAW 403

RESULT 9  
R77867  
ID R77867 standard; Protein; 408 AA.  
AC R77867;  
DT 13-NOV-1995 (first entry)  
DE S. clavuligerus ORF10 product.  
KM Clavulinate acid; clavulinate; antibiotic; beta-lactamase-inhibitor.  
OS Streptomyces clavuligerus.  
PN CA2108113-A.  
PD 09-APR-1995.  
PF 08-OCT-1993; 108113.  
PR 08-OCT-1993; CA-108113.  
PA (UYAL-) UNIV ALBERTA.  
PI Aidoo KA, Jensen SE, Paradkar AS;  
DR WPI: 95-207301/28.  
DR N-PSDB: Q91580.  
PT Clavulinate acid biosynthesis enzymes and corresp. DNA - useful for  
PT biosynthesis of the antibiotic in Streptomyces hosts which do not  
PT naturally produce clavulinate.  
PS Claim 32; Fig.19; 41pp; English.  
CC A 15 kb fragment S. clavuligerus NRRL 5741 genomic DNA (Q91580),  
CC extending downstream from pccC, included 10 ORFs encoding the  
CC enzymes required for clavulinate biosynthesis. The ORF10  
CC product (R77867) showed high similarity to cytochrome P450-type  
CC enzymes from other Streptomyces spp.  
SQ Sequence 408 AA;

Query Match 14.7%; Score 321; DB 1; Length 408;  
Best Local Similarity 28.8%; Pred. No. 9; 2e-25;

Matches 121; Conservative 59; Mismatches 186; Indels 54; Gaps 18;

QY 16 PNHPEHLPDFDMWNPNSLGSQVQEAAMAVIOESVNPVLWTRCGNH-WIATRQGLREA 74  
DB 14 PAVPMHRYCPVD---PPQLAGLSOKAASRYT-----LW---DGSVWLVTSAGARAV 62  
QY 75 YEDYRHFS-SECFIP-----REAGAYDFIPSPMDPEQORPA-----LANOV 118  
DB 63 LGDRFTAVYSAPGFPMILTRTSQLVYRANPESASTI--RMDDPQHSRLRSMTREFLARRA 120  
QY 119 VGM-PVVDKLENRIQELACSLIESLRPQOCNFTEDYAEPPPIRIFMLAGLPEDIPLH 177  
DB 121 EALPRVAREL---IDELTIGLVKGERP---VDLVAGLTIPPSRVITLLFGAGDREEFI 174  
QY 178 K-----YLTDOMTPRDSGMTFAEAKALYDILPIIEOROKPGTDAISIVANGOVNRP 233  
DB 175 EDRSAVLIDGTYPE---QVAKARDELJDGYLREYVERIEHPGDLISRLVIDQVRGHL 231  
QY 234 TSDEAKRMFGLLVGLDVTNVNLFSSMEFLAKSPERHQLERPERIPACCELLRRFS 293  
DB 232 RVEEMVPMCRULLVAGHTTSSQASLSLSLIDPELAGRTEDRPALLPRAVEELLRFHS 291  
QY 294 LVADG--RLITSDYEFHGVOLKKGQDILLPOMLSGLDERKNACPMDVDFSRQVSHTFG 351  
DB 292 IVQGLARAAVEDVQDLDVILRAGEGVYVLSAGNRDETLPDPDRVDVDRARHILAFG 351  
QY 352 HGSMLICGQHLAR--REIIVTKEMLRIPDSIA-PGAQIQKSGI-VGVALP 406  
DB 352 HGMHQCIGQWLARVELLEILAVALRW---PGARLAVPEFLDRHREVSIGCALPVTW 408

RESULT 10  
R38309  
ID R38309 standard; Protein; 412 AA.  
AC R38309;  
DT 04-DEC-1993 (first entry)  
DE Sequence of the P450-soy protein.  
KM P450soy; soy; gene; oxidation; haem protein.  
OS Streptomyces griseus ATCC 13273.  
PN W09312236-A.  
PD 24-JUN-1993.  
PF 16-DEC-1992; U10885.  
PR 16-DEC-1991; US-807001.  
PA (DUPO ) DU PONT DE NEMOURS & CO E.I.  
PI Omer CA, Sariaslani FS, Trower MK;  
DR WPI: 93-214178/26.  
DR N-PSDB: Q45569.  
PT Constitutive expression of P450 SOY (SOYC) and ferredoxin soy  
PT (soyD) in Streptomyces - used for oxidn. of organic chemicals  
PS Example: Figure 2; 45pp; English.  
CC Cytochrome P450soy was purified from S.griseus ATCC 13273. Two  
CC similar forms of P450soy were isolated. P450soy-delta, is derived  
CC from P450soy by in vitro proteolysis during isolation. One of the  
CC tryptic peptide fragments of cytochrome P450soy and of of the  
CC P450soy-delta protein were subjected to automated degradation. The  
CC NH2 terminal sequences are given in R38306 and R38307. A mixture of  
CC oligos that consist of possible DNA sequences that could encode the  
CC AAs FGWHQCL of the tryptic peptide was made. It consists of the  
CC sequences in Q43290-23. The oligo mixture was end-labeled and used  
CC to probe the EMBL4 library of S.griseus DNA. Hybridising plaques  
CC were isolated and a 4.8kb SacI DNA fragment was isolated from one  
CC clone that hybridised to the oligo probe mixture. As segment of the  
CC 4.8kb fragment was sequenced and found to contain an ORF. Within  
CC this ORF was a section that corresp. exactly to the AA sequence  
CC determined from the cytochrome P450soy tryptic peptide (see Q45569,  
CC R38309). The gene encoding the P450soy protein was called soyC. Five  
CC nucleotides downstream from the stop codon for soyC another ORF was  
CC identified. This ORF encodes an apparent ferredoxin-like protein.  
CC The gene was designated soyB and the protein ferredoxin-soy.  
SQ Sequence 412 AA;

Query Match 14.4%; Score 314; DB 1; Length 412;

	Best Local Similarity	27.2%	Pred. No. 4.9e-24;	
	Matches	84;	Conservative	56; Mismatches 163; Indels 6; Gaps 5.
QY	MDPPEOROFRALANOVGM	PVVDKLENNR	IOEACSLIESIR	POG-QCNFTEDYAEPPIR 161
	103			
Db	105	VDDPHNTORRRLITFES	YKRIGALRPRIQET	YVDLLDMEQPPAEIVSFALEPVSM 164
QY	162	IFMLAGLEPDEIPHLKYL	TDMPRDSMTFAEKA	EALYDLIRPIEORRQKPGTDAIS 221
	165	VICALGVPYADNAHAF	FEERSQRLRGPCADV	NRRARDELEFGLALIDKRAEPGGDILD 224
Db	222	IVANGOVNGRPITSD	EAKRMFGLLVGGLD	IVNFLFSMEFLANSPEHROELIERPERI 281
	225	ELIHRDHPDGPYDRQ	LVAFVILLIAGHET	TANNISLGTFTLLSHPEGLAALRRGGTST 284
QY	282	PACCELLRRSLVADG--	RITTSYEFHGVOLK	KGDQILLPOMLSGLDERNNACPMHVD 339
	285	AVVVEELL-RFLSL	IEGFLRLATEDME	VDGARIRGEGVVFSTSLINRDADVFRAETLD 343
QY	340	FSRQKVSHTTGSHGL	LCGLHARREITVIL	KEULTRIPDSIA-PCAQIOHKSG-IYS 397
	344	MDRPARRHHLAFG	FGVHOCIGOMLARA	EELDIAMRTLEFRLPGRLAVPAHEIRHKKGDTIQ 403
QY	398	GVQALPLVM 406		
	404	GLLDLPVAM 412		

RESULT 11  
R51368  
ID R51368 standard; Protein; 410 AA.  
AC R51368;  
DT 24-NOV-1994 (first entry)  
DE Protein containing Cytochrome P450 SCA-2 activity.  
KW Cytochrome P450 SCA-2; Streptomycetes carbophilus;  
KW treatment of hyperlipidaemia; drug preparation.  
OS Streptomycetes carbophilus.  
FH Key Location/Qualifiers  
FT protein 1..410  
FT /label= cytochrome P-450 containing activity  
PN J06070780-A.  
PD 15-MAR-1994.  
PF 28-AUG-1992; 229969.  
PR 28-AUG-1992; JP-229969.  
PA (SANTO) SANKYO CO LTD.  
DR WPI; 94-128679/16.  
DR N-PSDB; Q61452.  
PT Cytochrome P450 SCA-2 gene - from Streptomycetes carbophilus,  
PT useful in treatment of hyperlipidaemia  
PS Claim 1; Page 12-14; 18pp; Japanese.  
CC R51368 shows a protein having cytochrome P-450 activity. P-450  
CC SCA-2 can be prepared commercially for use in the preparation of a  
CC drug for the treatment of hyperlipidaemia.  
SQ Sequence 410 AA;

Query Match	14.38;	Score 311.5;	DB 1;	Length 410;
Best Local Similarity	25.68;	Pred. No. 8.9e-24;		
Matches 92; Conservative	63;	Mismatches 185;	Indels 19;	Gaps 7

```
QY      63 WIAHQGLIRAVDYD-----HFSSECFIP--REAGEADFTPTSMDDPEQOFR 112
        | : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      56 WWWKHEAARLLADPRISDRILADEPAPSPREKARGOSPAC---ICMDPEHCOTR 111
        | : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      113 ALANQVGMPLVDLKNRIEOLACLSELRPGS-QCNFEDAEAPEPINIFMLGLPE 171
        | : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      112 RMTISEFTVKRIKMRPDVERIYHGFTIDDLMLAGPTADVLSQEFLPVSWICHMGV 171
        | : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      172 EDPIHLKYLTFOOMRPDGSMTPFAEKALIVDLLPIFEORROKKGPAISIVANGOVN 231
        | : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      172 ADHFPODASRLVQAVDABSAVAARDFERITLDGLTTLKLESEFGTGLCKLVTHQADG 231
        | : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      232 PITSDEAKRMFGILLVGSLDTVVNFLSFSMEFLAKSPENHOELIERPERIPAACEELL 291
```

Db	232	EIDRELLSTALLNIVGHETTASMTSLSVTTLEHEDQNALRADSVLPGAVEELLRV	291
Qy	292	FSL--VADGRILTSDEYFHVOLKKGDOILLPOMLSGDERKNACPMHNVDFSRQVSHTT	349
Db	292	LAIADIAGRATADIETIDGOLLIRAGEVITVNSIANDSDSVFENPDRLDVHSNRHLS	351
Qy	350	FGHSHLCLGONLARRELIYVLEKMLIRIPDFSLA--FGAOLQIKSG--TVSGVALPLVW	406
Db	352	FGYVNOHCLGONLARLEVIYVLEFRLIRPRLRLAIVAEVQGLTLTRPGTITIGVLELPTW	410

RESULT 12  
R14724  
ID R14724 standard; Protein; 404 AA.  
AC R14724;  
DT 28-JAN-1992 (first entry)  
DE 6-hydroxylase encoded by EryF gene.  
KW C-6 hydroxylation; erythromycin; 6-deoxyerythromycin; antibiotics;  
KW Saccharopolyspora; erythromycin P450 monooxygenase; ss.  
OS Saccharopolyspora erythraea.  
PN MO9116334-A.  
PD 31-OCT-1991.  
PF 16-APR-1991; U02600.  
PR 18-APR-1990; US-510483.  
PA (ABBO ) ABBOTT LABORATORIES.  
PI Weber JM;  
DR WPI: 91-339744/46.  
DR N-PSDB; Q14548.  
PT New 6-deoxyerythromycin derive. - are antibiotics with increased  
PT acid stability, produced by cultivation of saccharopolyspora.  
PS Disclosure: Fig 3; 56pp; English.  
CC The eryF gene encodes the 6-hydroxylase component of the cytochrome  
CC P450 monooxygenase system responsible for the hydroxylation of 6-de-  
CC oxyerythronolide B to erythronolide B. Interruption of this step  
CC results in the formation of deoxyerythromycin A and new derivatives  
CC useful as antibiotic which have better stability against acids that  
CC the corresponding erythromycins. Interruption of the reaction can  
CC be effected by an insertion into the eryf gene of a plasmid, gene  
CC replacement or chemical or light-induced mutagenesis.  
C See also Q14549.  
SQ Sequence 404 AA;

Query Match 13.5%; Score 294.5; DB 1; Length 404;  
Best Local Similarity 24.2%; Pred. No. 4.9e-22;  
Matches 90; Conservative 74; Mismatches 173; Indels 35; Gaps 10;

QY	63	WIAIRGOLIRAVEDYDNHFS-----ECPFIIRAG---EAYDIPISM---DPPEQ	108
Db	40	WLWVGVEAKAALSIDLRISSDPKKRYGVEYEF--PAYLGEPEDEVIRFANNMGTSDPPTH	98
QY	109	ROFALANQVGMVGVYKLEIRIOELACSLIESLRPGGOCNFEDVNEPPPIRPMILAG	168
Db	99	TRLKLYSQEETTVARVEAMRPVQIITAEILIDEVGDGSDVYDIYRRNHPRLPIVICELG	158
QY	169	LPEEDIPHLKYLTD-----QMTRPDGSMTFAEAKALYDILPIIEOROKPGTDAI	220
Db	159	VDE-----KRGEGFRKSSILYMDPERABORQARREVNFIIDLVERRRTRPEGDLL	212
QY	221	SIVANG--VNGRPITSDAEAKRMGLLLVSGDLTVNFSLSMEFLAKSPENROELIERP	278
Db	213	SALIRVQDDDDGR-LSADELSSIALVLLAGFEASVSLIGIGYILLTLTHPDQALAVRDP	271
QY	279	ERIPACEELLIRRSI--VADGRILTSYEFHGVQALKGDDOILLPOMISGIDERKNACPMH	337
Db	272	SALFNAAVEEILRYTALOETTTREFAEDLEIGVAIPQSYTVILVANGCANMDPKQFPDPR	331
QY	338	VDEPRQKVSHTTECHGSHLICGOLARREIIVTLKELVTRIPDFSTAPGQ--IQKSGI	395
Db	332	SDVTRDTRGHSTSGGILHPCMGRLAKLEGEVALRLEGFRPALSLGIDADDVVMRRSVL	391
QY	396	VSGVQALPLVMD	407



KW Daunomycin C-14 hydroxylase; doxa gene; doxorubicin; daunomycin;  
KW 13-dihydrocarminomycin; carminomycin; anthracycline;  
KW anticancer; cytostatic; cancer; therapy; plasmid pANT199.  
OS Chimeric - streptomycetes sp. strain C5.  
OS Chimeric - synthetic.  
FH Key Location/Qualifiers  
FT Cleavage-site 31  
FT Protein /note="enterokinase cleavage site"  
FT 53..474  
PN W09744439-A2. /note="native daunomycin C-14 hydroxylase"  
PD 27-NOV-1997.  
PE 22-MAY-1997; U08690.  
PR 24-MAY-1996; US-653650.  
PA (OHIS ) UNIV OHIO STATE RES FOUND.  
PI Desanti CL, Dickens ML, Strohl WA;  
DR WPI: 98-018495/02.  
DR N-PSDB: V01452.  
PT Producing doxorubicin from daunomycin by daunomycin C14 hydroxylase  
PT - also hydroxylation and oxidation of other anthracycline(s) with  
PT the same enzyme  
PS Disclosure: Fig 9; 59pp; English.  
CC This protein comprises an N-terminal modified enzyme derived from  
CC the daunomycin C-14 hydroxylase (see W36128) of Streptomyces sp.  
CC strain C5. It is encoded by a gene construct (see V01452) in  
CC plasmid pANT199. In this plasmid, the doxa gene (see V01447) is  
CC translationally fused with a leader sequence encoding 6  
CC histidine residues so that the fusion protein can be affinity  
CC purified on a nickel-agarose gel. Daunomycin C-14 hydroxylase  
CC is a P450-like enzyme capable of converting daunomycin to the  
CC anticancer agent doxorubicin. Host cells, especially Streptomyces  
CC host cells, transformed with pANT199 can be used in methods for  
CC the production of doxorubicin from daunomycin or for the  
CC hydroxylation and oxidation of other anthracyclines.  
SQ Sequence 474 AA;

Query Match 7.4%; Score 161; DB 1; Length 474;  
Best Local Similarity 23.6%; Pred. No. 3.6e-08;  
Matches 96; Conservative 51; Mismatches 205; Indels 54; Gaps 14;

QY 39 QEAMAVIQESNVDPVTRNGGHMITRGQILREATEDYRHSSF--CPFIPEAGEAY 96  
DB 83 REAGPVV-EVNAP-----AGRAWITTDALAREVLADRFVKGPDLPATMNGVDDGL 135  
QY 97 D-----FIPSMPPPEOROFALANQVGMVVDKLENRIQELACSLIESL-----R 143  
DB 136 DIVPELRPTTLAVDGEDHRLRLRIHAPAFNFRRLAERTDRIATAIDRLITELADSSDR 195  
QY 144 PGGCQNFETDYAEFPPIRIFMLAGLP-----EEDIPHLKYL-----TDQMTRPD 188  
DB 196 SGEPAEILGGFAVHFPLVLICELGVPTDPAAREAVGVKALGLGQPSAGDGTDP 255  
QY 189 GSKTFEAKFALDYLPITTEORRQKGTDAISIVANGQVNGRPTSDKRMFGLLVG 248  
DB 256 GDVPDTSALTESL--LEAVNAARRKDRITRYLYERAQAEFGSVSDQLVYMITGLIFA 313  
QY 249 GLDPTVNFLSFMSFELAKSPENROELIERPERIPACCELLRR-----FSIVADGRILTS 303  
DB 314 GHDTGSFELG---LLAEVLAGRLADADGDAISRFEALRHHPVPSL---WRFAT 367  
QY 304 DYFEGVQLKKGQILLPQMLSGLDERKNACPMHVDKFSRQKVSHTTFFGSHLCLGQHILA 363  
DB 368 EVVIRGVRLPRGAPVLDIEGTMTDGRHNDAPAFHPRDRSRRLTFPGDPRHYCIGBOLA 427  
QY 364 RREIIVTKEMLRIPDFSLA--PGAQIQ--HKSGIVSGVALPLVW 406  
DB 428 QLESRTMIGVLRSRFPQARLAVPYEELRMCRCGAQTARLTDLR-VW 472



**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 12:38:07 ; Search time 66.92 Seconds  
(Without alignments)  
103.693 Million cell updates/sec

Title: US-09-246-451-13

Perfect score: 2177

Sequence: 1 TTERIQSNANLAPLPHPVE.....IVSGVQALPLVWDPATKAV 414

Scoring table: BLOSUM62  
Gapop 10.0 , Gapept 0.5

Searched: 164575 seqs, 16761186 residues

Total number of hits satisfying chosen parameters: 164575

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents-AA:\*  
1: /cgml\_7/prodata/1/1aa/5A.COMB.pep:\*  
2: /cgml\_7/prodata/1/1aa/5B.COMB.pep:\*  
3: /cgml\_7/prodata/1/1aa/5.COMB.pep:\*  
4: /cgml\_7/prodata/1/1aa/ECTUS.COMB.pep:\*  
5: /cgml\_7/prodata/1/1aa/backfilist1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	ID	Description
1	363.5	16.7	416	US-09-320-878-18	Sequence 18, Appl
2	340	15.6	403	5212296-9	Patent No. 5212296
3	328	15.1	406	5212296-6	Patent No. 5212296
4	314	14.4	412	US-08-102-863-11	Sequence 11, Appl
5	314	14.4	412	PCT-US92-10885-11	Sequence 11, Appl
6	271.5	12.5	419	US-09-335-409-8	Sequence 8, Appl
7	161	7.4	422	US-09-096-982-5	Sequence 5, Appl
8	161	7.4	422	US-08-653-650A-5	Sequence 5, Appl
9	161	7.4	474	US-09-096-982-8	Sequence 8, Appl
10	158	7.3	443	US-09-096-982-9	Sequence 9, Appl
11	158	7.3	443	US-08-653-650A-9	Sequence 9, Appl
12	158	7.3	443	US-08-396-218-2	Sequence 2, Appl
13	156	7.2	422	US-08-160-116-2	Sequence 2, Appl
14	156	7.2	422	US-08-194-981E-5	Sequence 5, Appl
15	131	6.0	512	US-09-320-878-7	Sequence 7, Appl
16	123.5	5.7	382	US-08-201-118-3	Sequence 3, Appl
17	116.5	5.4	490	US-08-201-118-9	Sequence 9, Appl
18	116.5	5.4	490	US-08-238-821B-3	Sequence 3, Appl
19	116.5	5.4	490	US-08-238-821B-9	Sequence 9, Appl
20	116.5	5.4	490	PCT-US95-05744-3	Sequence 3, Appl
21	116.5	5.4	490	PCT-US95-05744-9	Sequence 9, Appl
22	116.5	5.4	490	US-08-906-791-2	Sequence 2, Appl
23	113	5.2	480	US-08-201-118-7	Sequence 7, Appl
24	112.5	5.2	480	US-08-238-821B-7	Sequence 7, Appl
25	112.5	5.2	480	PCT-US95-05744-7	Sequence 7, Appl
26	111.5	5.1	504	US-08-457-274A-25	Sequence 25, Appl
27	111.5	5.1	504	PCT-US95-05758-25	Sequence 25, Appl
28	111.5	5.1	504	PCT-US95-05758-25	Sequence 25, Appl

29	108	5.0	492	3	US-08-724-466B-2	Sequence 2, Appl
30	107	4.9	513	3	US-08-948-564-6	Sequence 6, Appl
31	104.5	4.8	500	4	PCT-US95-13051-2	Sequence 2, Appl
32	104.5	4.8	500	4	PCT-US95-13051-2	Sequence 2, Appl
33	103.5	4.8	472	2	US-08-622-166A-2	Sequence 2, Appl
34	103.5	4.8	472	2	US-08-622-166A-4	Sequence 4, Appl
35	103	4.7	497	3	US-08-724-466B-4	Sequence 4, Appl
36	100.5	4.6	498	1	US-08-457-274A-24	Sequence 24, Appl
37	100.5	4.6	498	4	PCT-US95-05758-24	Sequence 24, Appl
38	99.5	4.6	480	2	US-08-201-118-1	Sequence 1, Appl
39	99.5	4.6	480	2	US-08-238-821B-1	Sequence 1, Appl
40	99.5	4.6	480	4	PCT-US95-05744-1	Sequence 1, Appl
41	98	4.5	490	1	US-08-201-118-5	Sequence 5, Appl
42	98	4.5	490	2	US-08-238-821B-5	Sequence 5, Appl
43	98	4.5	480	4	PCT-US95-05744-5	Sequence 5, Appl
44	96	4.4	480	1	US-08-201-118-11	Sequence 11, Appl
45	96	4.4	490	2	US-08-238-821B-11	Sequence 11, Appl

ALIGNMENTS

```
RESULT 1
US-09-320-878-18
; Sequence 18, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,558
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-320-878-18
```

Query Match 16.7%; Score 363.5; DB 3; Length 416;  
Best Local Similarity 28.3%; Pred. No. 2.1e-29;  
Matches 97; Conservative 62; Mismatches 163; Indels 21; Gaps 7;

QY	76	EDYHFSSECFPIREGEAYDFTPTSMDDPEQOFALANQVGMVYDKLEKRIODELA	135
DB	72	KDWR--NSTPPLTEAEALNNHMLE--DPRHRLKRLVAREFTMRVLELRPOVEIV	127
QY	136	CSLSESL--RPOGOCNTEDEYAEPPPIRIFMLLAGLPEEDIPHLKYLTDQWTRDGSMTF	193
DB	128	DGLVDALAPADGDADLMESLAWPLPITVISELLGVEBPDAARFVTDFAVFPPDDPAQA	187
QY	194	AEAKALYDYLPIPIEOROKPGTDAIS--IVANGQVNGRPTSDAKRMGLLVGGLDT	252
DB	188	QTMAKESGYLSRLIDSKRGDDGDLISALVTRTSDGDSRLTSELLGMHILLVACHET	247



```

; SOFTWARE: PatentIn Release #1.0,
; SOFTWARE: Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/102,863
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/807,001
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: GALLEGOS, R. THOMAS
; REGISTRATION NUMBER: 32,692
; REFERENCE/DOCKET NUMBER: CR-9000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-7342
; TELEFAX: 302-892-7949
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 412 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-102-863-11
```

Query Match 14.4% Score 314; DB 1; Length 412;

Best Local Similarity 27.2% Pred. No. 2.8e-24; Matches 84; Conservative 56; Mismatches 163; Indels 6; Gaps 5;

```

QY 103 MDPEQROFRALANOVVMPVVDKLENIQELACSLIESLRPOG-QCNFTDYAEPPIR 161
   :| | | | | : : : | | | | | : : : | | | : | | |
DB 105 VDDEHNTQRMMLPTPSVKRIGALRPRIQETVRLDAMERQGPRAELVSATLPPVSM 164
   :| | | | | : : : | | | | | : : : | | | : | | |
QY 162 IFMLLAGLPEDIDHLKYLTDQMTRPDGSMTFAEAKELVDYLPIITEOROKPGTDAIS 221
   :| | | | | : : : | | | | | : : : | | | : | | |
DB 165 VICALLGVPAADHAFEEERSQRLRGPGADVDNRADELLEYLGLALIDRKRAEPGDGLD 224
   :| | | | | : : : | | | | | : : : | | | : | | |
QY 222 IVANGOVNGRPTSDCAKRMFGILLVGLDITVNVFLFSMFLAKSPHRELIERPERI 281
   :| | | | | : : : | | | | | : : : | | | : | | |
DB 225 ELIHRDHPDGVDRQOLVAFVAILLIAGHETTANMISIGFTLLSHPEOLALRAGTST 284
   :| | | | | : : : | | | | | : : : | | | : | | |
QY 282 PAACEELLRRFSIADG--RLTSDYEFHGYOLKKGDQILLPQMSGLDERKNACPMHYD 339
   ||||| | | :| | | | | :| | | | | :| | | | | :| | |
DB 285 AVVVEELL-RLSTAEGQRLATEDMEVDGATIRKGEVVESTSLINDADVPRAEITLD 343
   :| | | | | :| | | | | :| | | | | :| | | | | :| | |
QY 340 FSRQKVSHTTFFGHSHLCLGHLARREIIVTLKEMLTRIPDFSIA-PCAQIQHKSQ-IVS 397
   :| | | | | :| | | | | :| | | | | :| | | | | :| | |
DB 344 WDRPARHHLAFGFCVHOCIGNLARAEILDIAMTLFEHLPGRLAVRAHEIRHKPGDTIQ 403
   :| | | | | :| | | | | :| | | | | :| | | | | :| | |
QY 398 GVQALPLVW 406
   :| | | | | :| | | | | :| | | | | :| | | | | :| | |
DB 404 GLDLPLVAM 412
```

RESULT 5  
PCT-US92-10885-11

```

; Sequence 11, Application PC/TUS9210885
; GENERAL INFORMATION:
; APPLICANT: SARIASLANI, SIMA
; TITLE OF INVENTION: CONSTITUTIVE
; TITLE OF INVENTION: EXPRESSION OF P450SOY
; TITLE OF INVENTION: AND FERREDOXIN-SOY IN
; TITLE OF INVENTION: STREPTOMYCES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS
; ADDRESSEE: AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: USA
; ZIP: 19898
```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch,
; MEDIUM TYPE: 1.0 MB
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh System, 6.0
; SOFTWARE: Microsoft Word, 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/10885
; FILING DATE: 19921216
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: GALLEGOS, R. THOMAS
; REGISTRATION NUMBER: 32,692
; REFERENCE/DOCKET NUMBER: CR-9000-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-7342
; TELEFAX: 302-892-7949
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 412 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; PCT-US92-10885-11
```

Query Match 14.4% Score 314; DB 4; Length 412;

Best Local Similarity 27.2% Pred. No. 2.8e-24; Matches 84; Conservative 56; Mismatches 163; Indels 6; Gaps 5;

```

QY 103 MDPEQROFRALANOVVMPVVDKLENIQELACSLIESLRPOG-QCNFTDYAEPPIR 161
   :| | | | | : : : | | | | | : : : | | | : | | |
DB 105 VDDEHNTQRMMLPTPSVKRIGALRPRIQETVRLDAMERQGPRAELVSATLPPVSM 164
   :| | | | | : : : | | | | | : : : | | | : | | |
QY 162 IFMLLAGLPEDIDHLKYLTDQMTRPDGSMTFAEAKELVDYLPIITEOROKPGTDAIS 221
   :| | | | | : : : | | | | | : : : | | | : | | |
DB 165 VICALLGVPAADHAFEEERSQRLRGPGADVDNRADELLEYLGLALIDRKRAEPGDGLD 224
   :| | | | | : : : | | | | | : : : | | | : | | |
QY 222 IVANGOVNGRPTSDCAKRMFGILLVGLDITVNVFLFSMFLAKSPHRELIERPERI 281
   :| | | | | : : : | | | | | : : : | | | : | | |
DB 225 ELIHRDHPDGVDRQOLVAFVAILLIAGHETTANMISIGFTLLSHPEOLALRAGTST 284
   :| | | | | : : : | | | | | : : : | | | : | | |
QY 282 PAACEELLRRFSIADG--RLTSDYEFHGYOLKKGDQILLPQMSGLDERKNACPMHYD 339
   ||||| | | :| | | | | :| | | | | :| | | | | :| | |
DB 285 AVVVEELL-RLSTAEGQRLATEDMEVDGATIRKGEVVESTSLINDADVPRAEITLD 343
   :| | | | | :| | | | | :| | | | | :| | | | | :| | |
QY 340 FSRQKVSHTTFFGHSHLCLGHLARREIIVTLKEMLTRIPDFSIA-PCAQIQHKSQ-IVS 397
   :| | | | | :| | | | | :| | | | | :| | | | | :| | |
DB 344 WDRPARHHLAFGFCVHOCIGNLARAEILDIAMTLFEHLPGRLAVRAHEIRHKPGDTIQ 403
   :| | | | | :| | | | | :| | | | | :| | | | | :| | |
QY 398 GVQALPLVW 406
   :| | | | | :| | | | | :| | | | | :| | | | | :| | |
DB 404 GLDLPLVAM 412
```

RESULT 6  
US-09-335-409-8

```

; Sequence 8, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
```

SEQ ID NO 8  
LENGTH: 419  
TYPE: PRT  
ORGANISM: Sorangium cellulosum  
US-09-335-409-8

Query Match 12.5%; Score 271.5; DB 3; Length 419;  
Best Local Similarity 23.8%; Pred. No. 7.4e-20;  
Matches 100; Conservative 67; Mismatches 179; Indels 75; Gaps 14;

QY 6 OSNANLAPLPHVPEHVLVPEEDMNPNSLSAGVQ-----AMAVLQESNVDLVWTRKNG 61  
DB 3 QEOANQSETPR-----AFDKPPAP-----GVAEDPPALIERLEA-TPIFYWD--EGR 48  
QY 62 HWIATRGQ-----LIREAYEDYRHFSECEPIFREAGEAYDFTPTSMDPPEQR 109  
DB 49 SWVLTRHDVSAVFRDRFRVSVSREMESSAYSNAIP-----ELSDMKKYGIFGLRPEDHA 104  
QY 110 QFRALANOVGMVYDKLENRIQELACSLIESLRPGQCNFTEDYAEFPPIRIFMLIAGL 169  
DB 105 RVKRLVNFSTSRALIDLRRAEIQRTVQDLIDARSGQEEFDVVRDYAGIIPMAISALLKV 164  
QY 170 PEEDIPHLKTLTQMTPRDSMTFAEKEALYDILP-----I 207  
DB 165 PAE-----CDEKFRFESSAT---ARALGVGLVQVDEETKTIVASYTEGALLHDV 212  
QY 208 IEOROKP-GTDAISIVANGOVNGRPITSDEAKRMFGLLVGLDVVNFSLFSMEFLAK 266  
DB 213 LDERRRRLPNDVLTMLQAEADSGRLSTKEVALVGAITIAAGDTITIIYLAIRVNLRL 272  
QY 267 SPEHQELIERPERIPACEELLRFSLVADG--RILTSDEYHGVOLKKGDQI--LLPQ 322  
DB 273 SPEALELVKAEPLMRNALDEVLEFDNLIRGVRFARQDEYCGASIKKGEWVFLIPS 332  
QY 323 MLSGLDERKNACPMHVDFSQKVSHTFGSHCLGQHLARRIITLKEMLTRIDFS 382  
DB 333 ALR--DGVTSRPDVFVDRBDTGSALVGRGPHVCPSLARLEAEIAGVITFRPEMK 390  
QY 383 I 383  
DB 391 L 391

RESULT 7  
US-09-982-5  
Sequence 5, Application US/09096982  
Patent No. 5962293  
GENERAL INFORMATION:  
APPLICANT: Strohl, William R.  
APPLICANT: Dickens, Michael L.  
APPLICANT: Desanti, Charles L.  
TITLE OF INVENTION: METHOD OF PRODUCING DOXORUBICIN  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CALFEE, HALTER & GRISWOLD  
STREET: 800 Superior Avenue, Suite 1400  
CITY: Cleveland  
STATE: Ohio  
COUNTRY: USA  
ZIP: 44114-2688  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/096,982  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Goltick, Mary E.  
REGISTRATION NUMBER: 34829

REFERENCE/DOCKET NUMBER: 22727/00131  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 216-622-8458  
TELEFAX: 216-241-0816  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 422 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-982-5

Query Match 7.4%; Score 161; DB 2; Length 422;  
Best Local Similarity 23.6%; Pred. No. 2.2e-08;  
Matches 96; Conservative 51; Mismatches 205; Indels 54; Gaps 14;

QY 39 QEAMVQLQESNVDLVWTRKNGHWITRQQLREAYEDYRHFSS--CFPIFREAGEAY 96  
DB 31 REAGPVV-EVNAP-----AGRAWVITDDALAREVLADRFVKGPDLAFTAMRGVDDGL 83  
QY 97 D-----FIPSMDEPQORFALANOVGMVYDKLENRIQELACSLIESL-----R 143  
DB 84 DIPPELPEPTLLAVDGEDHRLRLRIHAPRNPRLAERTDRIALADRLTLELADSSDR 143  
QY 144 PQGQCNFTEDYAEFPPIRIFMLIAGL-----EEDIPHLKYL-----TDQMTRPD 188  
DB 144 SGEPALIGFAFNAHFPFLVYCELLGVPTPRAMAREAVGLKALGCGPQSGAGDGTDPRA 203  
QY 189 GSNTPFAKALYDILPIETQRORQKPGTDAISIVANGOVNGRPITSDEAKRMFGLLVG 248  
DB 204 GDVPDTSALSL--LLEAVAHARRKDTRTMTRVLYERAQAEFSVSDDQLVYMITGLIFA 261  
QY 249 GLTPVNFSLFSMEFLAKSPENHQELIERPERIPACEELLAR-----PSLVADRLITS 303  
DB 262 GHDTGSFGLF---LAEVLAGRIADADGDAISREVEEALRIHNPVYSL---WRPAAT 315  
QY 304 DYFHGVOLKKGQDILPQMLSGIDERKNACPMHVDFSRQKVSHTFGSHCLGQHLA 363  
DB 316 EVVIRGVRLPRGARVLVDIEGTMTDGRHNDARPHAFHNDPRSRLRLTFGDSPHICIGOLA 375  
QY 364 RREIYTLKEMLTRIDFESTA-PSAQIO--HKSIVSGVALPLVW 406  
DB 376 QLESRTWIGVLRSPQARLAVPEELRMCCKGAGTARTLDP--VW 420

RESULT 8  
US-08-653-650A-5  
Sequence 5, Application US/08653650A  
Patent No. 5976830  
GENERAL INFORMATION:  
APPLICANT: Strohl, William R.  
APPLICANT: Dickens, Michael L.  
APPLICANT: Desanti, Charles L.  
TITLE OF INVENTION: METHOD OF PRODUCING DOXORUBICIN  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CALFEE, HALTER & GRISWOLD  
STREET: 800 Superior Avenue, Suite 1400  
CITY: Cleveland  
STATE: Ohio  
COUNTRY: USA  
ZIP: 44114-2688  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/653,650A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:



```
/ APPLICATION NUMBER: US/08/653,650A
/ FILING DATE:
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Golrick, Mary E.
/ REGISTRATION NUMBER: 34829
/ REFERENCE/DOCKET NUMBER: 22727/00131
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 216-622-8458
/ TELEFAX: 216-241-0816
/ INFORMATION FOR SEQ ID NO: 8:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 474 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-653-650A-8

Query Match 7.4%; Score 161; DB 2; Length 474;
Best Local Similarity 23.6%; Pred. No. 2.6e-08;
Matches 96; Conservative 51; Mismatches 205; Indels 54; Gaps 14;

QY 39 QEAMAVLQESNVDPDLVWTRCGHWTATRGQLIREAEDYRHSSE--CPPIRAGEAY 96
DB 83 REAGPVV-EVNAP-----AGGPAMVITDDALAREVLADRFYKGPDLATPAMRGVDDGL 135
QY 97 D-----FTPTMDPREQROFALANQVGMVVDKLENRIQELACSLIESL-----R 143
DB 136 DIVPELRPTLLAVDGEDHRLRIHAPAFNPRRLAERTDRIATAIDRLTELADSSDR 195
QY 144 PQGCNFTEDYAEPRPIRIFMLAGP-----EEDIPHLKYL-----TDQMTRPD 188
DB 196 SGPEAELIGGFAYHFPFLVLCGLGVPTDPMAREAVGLKALGLGPGSAGDGTDP 255
QY 189 GSMTFEAKKALDYLIPIEORRQKPGTDAISIVANGVNGRPITSDKRRFGILLVG 248
DB 256 GDVPDTSALSL--LLEAVIAARKKDTRTMTRVLYERAQAEFGVSDDOLVYMITGLIFA 313
QY 249 GLDTVNVFLFSMEFLAKSPENROELIERPERIPAAACEILLRR-----FSLVADGRILTS 303
DB 314 GHDTGTSFLGF--LLAEVLAGRLADADDDAISRFVEELRNHPVPYSL--WRAAT 367
QY 304 DYEPHGVOLKKGOILLPQMLSGDERKNACPMHVDPSRQKVSHTTFGSHLCLGQHILA 363
DB 368 EYVIRGVRLPRGAPVLDIEGTNTDGRHNDAPAFHPRDRSRRLTFGGDPHYCIGOLA 427
QY 364 RRELIYTLKEMLRIPDFSIA--PGAQIQ--HKSGLVSGVALPLVW 406
DB 428 QLESRTWIGVLRSPQARLAVPYEELRWCKGKAQTARLTDLR-VW 472

RESULT 11
US-09-096-982-9
/ Sequence 9, Application US/09096982
/ Patent No. 5962293
/ GENERAL INFORMATION:
/ APPLICANT: Strohl, William R.
/ APPLICANT: Dickens, Michael L.
/ APPLICANT: Desanti, Charles L.
/ TITLE OF INVENTION: METHOD OF PRODUCING DOXORUBICIN
/ NUMBER OF SEQUENCES: 9
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: CALFEY, HALTER & GRISWOLD
/ STREET: 800 Superior Avenue, Suite 1400
/ CITY: Cleveland
/ STATE: Ohio
/ COUNTRY: USA
/ ZIP: 44114-2688
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/096,982
/ FILING DATE:
/ CLASSIFICATION:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Golrick, Mary E.
/ REGISTRATION NUMBER: 34829
/ REFERENCE/DOCKET NUMBER: 22727/00131
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 216-622-8458
/ TELEFAX: 216-241-0816
/ INFORMATION FOR SEQ ID NO: 9:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 443 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
/ US-09-096-982-9

Query Match 7.3%; Score 158; DB 2; Length 443;
Best Local Similarity 23.6%; Pred. No. 4.8e-08;
Matches 95; Conservative 50; Mismatches 210; Indels 48; Gaps 13;

QY 39 QEAMAVLQESNVDPDLVWTRCGHWTATRGQLIREAEDYRHSSE--CPPIRAGEAY 96
DB 52 REAGPVV-EVNAP-----AGGPAMVITDDALAREVLADRFYKGPDLATPAMRGVDDGL 104
QY 97 D-----FTPTMDPREQROFALANQVGMVVDKLENRIQELACSLIESL-----R 143
DB 105 DIVPELRPTLLAVDGEDHRLRIHAPAFNPRRLAERTDRIATAIDRLTELADSSDR 164
QY 144 PQGCNFTEDYAEPRPIRIFMLAGP-----EEDIPHLKYL-----TDQMTRPD 188
DB 165 SGPEAELIGGFAYHFPFLVLCGLGVPTDPMAREAVGLKALGLGPGSAGDGTDP 224
QY 189 GSMTFEAKKALDYLIPIEORRQKPGTDAISIVANGVNGRPITSDKRRFGILLVG 248
DB 225 GDVPDTSALSL--LLEAVIAARKKDTRTMTRVLYERAQAEFGVSDDOLVYMITGLIFA 282
QY 249 GLDTVNVFLFSMEFLAKSPENROELIERPERIPAAACEILLRRSLV--ADGRILTS DYE 306
DB 283 GHDTGTSFLGF--LLAEVLAGRLADADDDAISRFVEELRNHPVPYSLWRAATEV 339
QY 307 FHGVOLKKGOILLPQMLSGDERKNACPMHVDPSRQKVSHTTFGSHLCLGQHILARE 366
DB 340 IRGVRLPRGAPVLDIEGTNTDGRHNDAPAFHPRDRSRRLTFGGDPHYCIGOLA 399
QY 367 IIVTLKEMLRIPDFSIA--PGAQIQ--HKSGLVSGVALPLVW 406
DB 400 SRIMIGVLRSPQARLAVPYEELRWCKGKAQTARLTDLR-VW 441

RESULT 12
US-08-653-650A-9
/ Sequence 9, Application US/08653650A
/ Patent No. 5976830
/ GENERAL INFORMATION:
/ APPLICANT: Strohl, William R.
/ APPLICANT: Dickens, Michael L.
/ APPLICANT: Desanti, Charles L.
/ TITLE OF INVENTION: METHOD OF PRODUCING DOXORUBICIN
/ NUMBER OF SEQUENCES: 9
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: CALFEY, HALTER & GRISWOLD
/ STREET: 800 Superior Avenue, Suite 1400
/ CITY: Cleveland
/ STATE: Ohio
/ COUNTRY: USA
/ ZIP: 44114-2688
/ COMPUTER READABLE FORM:
```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/653,650A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Goitlick, Mary E.
; REGISTRATION NUMBER: 34829
; REFERENCE/DOCKET NUMBER: 22727/00131
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 216-622-8458
; TELEFAX: 216-241-0816
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 443 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-653-650A-9
```

Query Match 7.3%; Score 158; DB 2; Length 443;

Best Local Similarity 23.6%; Pred. No. 4.8e-08; Mismatches 210; Indels 48; Gaps 13;

```

; 39 QEAMAVLQESNVDPDLVWTRCNGHMIAITRCQLIKEAYEDYRHFSS--CEPIPREAGEAY 96
; 52 REAGPVV-EVNAP-----AGRAWVITDDALAREVLADRFVVDPLATPAMRGVDGL 104
; 97 D-----FIPSMDEPQOFRALANQVGMVVDKLENRIQELACSLIESL-----R 143
; 105 DIPVPELRPTLLAVDGEDHRRRLRIHAPAFNPRRLAERTDRIAIADRLITELADSSDR 164
; 144 PQGQCNTEYAEPRIRIMLAGLP-----EEDIPHLKYL-----TDQMRPD 188
; 165 SGEPAELIGFAVHPLVLCGLGVPTDPMAREAVGLKALGLGSGPAGSGGDTPA 224
; 189 GSMFAFAKALDYLLPIEORROKPGTDAISIVANGVNGRPITSDEAKRMFG 248
; 225 GQVDTALSL--LLEAVNAARKKDTMTKRVYERAOAFGVSDDQLYMTGLIFA 282
; 249 GLDTVNFVFSMEFLAKSPENROELIERPERIPACEELLRFSLV--ADGRILTSDE 306
; 283 GHDTGSFLGF--LLEAVLAGRLAADADGDAISRVEEALRNHPVPTLMRPAATEV 339
; 307 FHCYQKLKKGQOILLQMLSGLDERKNACPMHVDPSROKVSHTTGHSHLCLGHLARRE 366
; 340 IRGRLPRGAPVLVDIEGTITDGRHNDAPHAFHEDRPSRRRLTFGGDPHYCIGOLAOLE 399
; 367 IIVLKEMLTRIPDFSIA-PGAQIQ--HKSGIVSGVQALPLVW 406
; 400 SRPMIGVLRSPQARLAVPYEELRMCKRKGAGTARLTDLR-VW 441
Db
```

```

RESULT 13
US-08-396-218-2
; Sequence 2, Application US/08396218
; Patent No. 5695966
; GENERAL INFORMATION:
; APPLICANT: INVENTI, Augusto
; APPLICANT: BREME, Umberto
; APPLICANT: COLOMBO, Anna L.
; APPLICANT: HUTCHINSON, Charles R
; APPLICANT: OTTEN, Sharee
; TITLE OF INVENTION: PROCESS FOR PREPARING DOXORUBICIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIKAIIDO, MARCELSTEIN, MURRAY & ORAM
```

```

; STREET: 655 Fifteenth Street, N. W., Suite 330 - G
; STREET: Street lobby
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-5701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/396,218
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: KITTS, Monica C.
; REGISTRATION NUMBER: 36,105
; REFERENCE/DOCKET NUMBER: P1615-5002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202/638-5000
; TELEFAX: 202/638-4610
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 422 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-396-218-2
```

Query Match 7.2%; Score 156; DB 1; Length 422;

Best Local Similarity 23.8%; Pred. No. 7.2e-08; Mismatches 202; Indels 58; Gaps 15;

```

; 39 QEAMAVLQESNVDPDLVWTRCNGHMIAITRCQLIKEAYEDYRHFSSCEPPIPREAGEAYDF 98
; 31 REAGPVV-EVNAP-----AGRAWVITDDALAREVLADRFVVD--PDLAPAMRG 78
; 99 IPTSM--PEQOFRALA-----NQVGMVVD---KLENRIQELACSLIESL- 142
; 79 VDDGLDIPVPELRPTLLAVDGEAHRRRLRIHAPAFNPRRLAERTDRIAIAGLLTELA 138
; 143 -----RPGQCNTEYAEPRIRIMLAGLP-----EEDIPHLKYL-----TDQ 183
; 139 DASGRSKPAELIGFAVHPLVLCGLGVPTDPMAREAVGLKALGLGSGPAGSGGD 198
; 184 MTRPDGSMTEFAKEALDYLLPIEORROKPGTDAISIVANGVNGRPITSDEAKRMFG 243
; 199 GTPDAGVPTDSALSL--LLEAVNSARKNDPTMTKRVYERAOAFGVSDDQLYMTIT 256
; 244 LLVGLDITVNFVFSMEFLAKSPENROELIERPERIPACEELLRFSLV--ADGRIL 301
; 257 GLIPAGHDITGSFLGF--LLEAVLAGRLAADADEDAVSREVEALRNHPVPTLMRFA 313
; 302 TSDYERHGYQKLKKGQOILLQMLSGLDERKNACPMHVDPSROKVSHTTGHSHLCLGQH 361
; 314 ATEVTIGVLRPRGAPVLVDIEGTITDGRHNDAPHAFHEDRPSRRRLTFGGDPHYCIGEQ 373
; 362 LAREIIVTLKEMLTRIPDFSIA-PGAQIQ--HKSGIVSGVQALPLVW 406
; 374 LAOLESRMTIGVLRSPPEARLAVPYDELRMCKRKGAGTARLTDLR-VW 420
Db
```

```

RESULT 14
US-08-760-116-2
; Sequence 2, Application US/08760116
; Patent No. 5786190
; GENERAL INFORMATION:
; APPLICANT: INVENTI, Augusto
; APPLICANT: BREME, Umberto
; APPLICANT: COLOMBO, Anna L.
; APPLICANT: HUTCHINSON, Charles R
```





Db 365 PYMEAFILLETFRHSSFVPTIPHSTRDYSLKGFYIPKGRCVFNOMQINHDOKLWVNP 424  
QY 336 -----MHVDSRQY---SHTEGHSHLCLGOHLARREIYTLKEMLTRIPDPSTAP 385  
Db 425 EFLPERFLTPDGAIDKVLSEKVIIFGMRKRCIGETVARWEVFLFLAILLQRY-EFSYPL 483  
QY 386 GAQI 389  
Db 484 GVKV 487

Search completed: October 4, 2000, 12:38:09  
Job time: 21426 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 12:36:43 ; Search time 73.09 Seconds  
(Without alignments)  
350.531 Million cell updates/sec

Title: US-09-246-451-13  
Perfect score: 2177  
Sequence: 1 TTTETIQSNANLAPLPHPVPE.....IVSGVOALPLWDPATTKAV 414

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 178050 seqs, 61884766 residues

Total number of hits satisfying chosen parameters: 178050

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: PIR\_64:\*  
2: PIR1:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2165	99.4	415	1	04PSCP
2	376	17.3	393	1	C70829
3	371	17.0	410	2	S39924
4	362	16.6	397	1	S51594
5	359	16.5	398	1	H70807
6	355	16.3	396	1	B69851
7	346.5	15.9	428	1	E70729
8	344.5	15.8	404	2	T30231
9	340	15.6	403	2	B33401
10	337	15.5	411	1	JC5859
11	334.5	15.4	410	1	O4B86M
12	334	15.3	410	1	E69611
13	333.5	15.3	405	1	B42606
14	328	15.1	406	2	A35401
15	327	15.0	406	1	A48495
16	325.5	15.0	376	1	G69679
17	323	14.8	428	1	A42971
18	318.5	14.6	405	1	H70752
19	317.5	14.6	399	1	A55578
20	314	14.4	412	2	S24750
21	311.5	14.3	410	2	JC4287
22	307	14.1	310	2	T44767
23	305.5	14.0	395	1	G69594
24	303	13.9	417	1	S49051
25	302	13.9	438	1	E70515
26	301	13.8	402	2	JC5151
27	301	13.8	404	1	JC5150
28	294.5	13.5	408	2	PD0007
29	294.5	13.5	408	2	JC5674

30	291.5	13.4	406	1	S18531	cytochrome P450 er
31	291.5	13.4	433	1	B70677	cytochrome P450 rv
32	285	13.1	381	1	S15809	cytochrome P450 cy
33	283.5	13.0	337	2	T30601	cytochrome P450 hy
34	273	12.5	414	1	E70708	cytochrome P450 rv
35	269.5	12.4	386	2	T30235	cytochrome P450 -
36	266.5	12.2	310	2	T44857	probable hydroxyla
37	265	12.2	412	1	B40634	erythromycin monoo
38	251.5	11.6	401	2	T36526	probable cytochrom
39	251	11.5	401	1	I40208	cytochrome P450 bt
40	250	11.5	411	1	JC4003	cytochrome P450 -
41	244	11.2	406	3	T17487	cytochrome P450 hy
42	240	11.0	433	2	T44587	cytochrome P450 ho
43	237	10.9	402	1	A70707	cytochrome P450 rv
44	235	10.8	396	1	H70730	cytochrome P450 rv
45	225	10.3	398	1	S47520	vitamin D-3 25-hyd

## ALIGNMENTS

RESULT 1  
04PSCP  
camphor 5-monooxygenase (EC 1.14.15.1) cytochrome P450 101 - Pseudomonas putida plasm  
N:Alternate names: cytochrome P450-CAM  
C:Species: Pseudomonas putida  
C:Date: 30-Apr-1982 #sequence.revision 31-Dec-1993 #text\_change 03-Mar-2000  
R:Accession: A25660; S34614; C60886; A00194  
R:Unger, B.P.; Gunsalus, I.C.; Sliagar, S.G.  
J. Biol. Chem. 261, 1158-1163, 1986  
A:Title: Nucleotide sequence of the Pseudomonas putida cytochrome P-450-cam gene and  
A:Reference number: A94678; MUID:86111751  
A:Accession: A25660  
A:Molecule type: DNA  
A:Residues: 1-415 <DNC>  
A:Cross-references: GB:M12546; NID:g151114; PIDN:AAA25760.1; PID:g151115  
R:Atamaki, H.; Koga, H.; Sagara, Y.; Hosoi, M.; Horiuchi, T.  
Biochim. Biophys. Acta 1174, 91-94, 1993  
A:Title: Complete nucleotide sequence of the 5'-exo-hydroxycamphor dehydrogenase gene  
A:Reference number: S34613; MUID:93326643  
A:Accession: S34614  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-42 <ARA>  
A:Experimental source: ppg3; ATCC 17453; CAM plasmid  
R:Romeo, C.; Moriwaki, N.; Yasunobu, K.T.; Gunsalus, I.C.; Koga, H.  
J. Protein Chem. 6, 253-261, 1987  
A:Title: Identification of the coding region for the putidaredoxin reductase gene fro  
A:Reference number: A60886  
A:Accession: C60886  
A:Molecule type: DNA  
A:Residues: 408-415 <ROM>  
R:Hanlu, M.; Ames, L.G.; Yasunobu, K.T.; Shastri, B.A.; Gunsalus, I.C.  
J. Biol. Chem. 257, 12664-12671, 1982  
A:Title: Amino acid sequence of the Pseudomonas putida cytochrome P-450. II. Cyanogen  
A:Reference number: A00194; MUID:83030788  
A:Accession: A00194  
A:Molecule type: protein  
A:Residues: 2-55,58-276,'Q',278-361,'S',363-407,'N',409-415 <HAN>  
C:Genetics:  
A:Gene: camc; CYP101  
A:Genome: plasmid  
C:Function:  
A:Description: catalyzes hydroxylation of camphor to yield 5-exo-hydroxycamphor; elac  
C:Superfamily: Pseudomonas plasmid camphor 5-monooxygenase; cytochrome P450 homology  
C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygena  
F:246-380/Domain: cytochrome P450 homology <CYP>  
F:358/Binding site: heme iron (Cys) (axial ligand) #status experimental

Query Match 99.4%; Score 2165; DB 1; Length 415;  
Best Local Similarity 99.5%; Pred. No. 3.6e-157;  
Matches 412; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY	1	TTESSNANLAPLPPHPEHLVPEDMYNPSNLSAGVOEAMAYLOESNVDLWTRNG	60
Db	2	TTEIIOANANLAPLPPHPEHLVPEDMYNPSNLSAGVOEAMAYLOESNVDLWTRNG	61
OY	61	GHWIATROQLIREAEEDYRHFSSSECPFIPIREAGEAYDIPIISMOPPEQORALANOVG	120
Db	62	GHWIATROQLIREAEEDYRHFSSSECPFIPIREAGEAYDIPIISMOPPEQORALANOVG	121
OY	121	MPVVDKLENRIQELACSLIESIRPOGOCNFTEDYAEPPRIIFMLLAGLPEDIDPHLYL	180
Db	122	MPVVDKLENRIQELACSLIESIRPOGOCNFTEDYAEPPRIIFMLLAGLPEDIDPHLYL	181
OY	181	TDQMRPPGSMYFAAKALYDYLPIIEORRQKPGTALISIVANGOVNGRPISDEKKR	240
Db	182	TDQMRPPGSMYFAAKALYDYLPIIEORRQKPGTALISIVANGOVNGRPISDEKKR	241
OY	241	MFGLLVAGGLTVVNFLEFSMEFLAKSPENHQELERPERIPACEEYLRFSTYADGRI	300
Db	242	MCGLLVAGGLTVVNFLEFSMEFLAKSPENHQELERPERIPACEEYLRFSTYADGRI	301
OY	301	LTSDEYFHGVOLKKSDQILLPQMLSGDDEKRNACPMHYDFSROKVSHTTFGHGSHLCGQ	360
Db	302	LTSDEYFHGVOLKKSDQILLPQMLSGDDEKRNACPMHYDFSROKVSHTTFGHGSHLCGQ	361
OY	361	HLAREELIVLAKEMLTRIPDSIAPGAOIOHKSIGVSGVQALPLVMDPATTKAV	414
Db	362	HLAREELIVLAKEMLTRIPDSIAPGAOIOHKSIGVSGVQALPLVMDPATTKAV	415

RESULT 2  
C70929  
cytochrome P450 RV1785c - Mycobacterium tuberculosis (strain H37RV)  
N:Contains: oxidoreductase (EC 1.-.-.-)  
C:Species: Mycobacterium tuberculosis  
C:date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: C70929  
R:Coile, S.T.; Brosch, R.; Parthill, J.; Garner, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajadream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987  
A:Accession: C70929  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-393 <COI>  
A:Cross-references: GB:AL022021; GB:AL123456; NID:3250699; PIDN:CAAI7707.1; PID:e125459  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: RV1785c  
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology  
C:Keywords: oxidoreductase

	Query Match	17.3%	Score 376;	DB 1;	Length 393;	
	Best Local Similarity	28.7%;	Pred.No. 3.8e-21;			
	Matches 110; Conservative	70;	Mismatches 183;	Indels 20;	Gaps 10;	
OY	31	PSNTSAGVOEAMAVLOESNVDPDLVWTRCNGCHWITRGQLIREAYEDRHFSSECPFIPR	90			
		:     :     :     :     :     :     :     :     :				
Dd	23	PMAVDRCGV--GKMTLRADG--PYVF--MNGWYIL-TREDVLALNRKNKYSSSKRALD-	74			
OY	91	EAGEAYDFIPTSMDPPEQROFRFLANOVGMVPYDKLENRIQELACSLIESLRPOGCNCF	150			
		: :       :     :     :     :     :     :     :				
Dd	75	-PGNPPLPVAPLAFDPPEHTRRYRLIQPFSPALASKALPSIRRHVAAMDIAAGRGCEEA	133			
OY	151	TEDVNEPPPIREIFEMTLASTPEDRIPHLTLYLDQM----TPDGSGTFEAKAEALTYDIIP	206			
		: :       :     :     :     :     :     :     :				
Dd	134	MAOLNLNFPEFOLEFLYLGLPLEDRDLIGMKDAVIAMSDDRHPPTEADYAARRELLEYTLTA	193			
OY	207	IIEGRQRKPGTDALSIIVANGCVNGRPITSDEAKRMFGILLVGGLDTVNNLSFSMEFLAK	266			

[illegible]

RESULT 3  
S39924  
cytochrome P450meg - *Bacillus megaterium* (ATCC 13368)  
C:Species: *Bacillus megaterium*  
A:Variety: ATCC 13368  
C:Date: 07-Oct-1994 #sequence\_revision 23-Feb-1996 #text\_change 04-Mar-2000  
C:Accession: S39924 #S32216  
R:Rauschenbach, R.; Isernhagen, M.; Noeske-Jungblut, C.; Boisdol, W.; Stewert, G.  
MOL. Gen. Genet. 241, 170-176, 1993  
A:Title: Cloning sequencing and expression of the gene for cytochrome P450meg, the str  
A:Reference number: S39924; MUID:94045677  
A:Accession: S39924  
A:Molecule type: DNA  
A:Residues: 1-410 <RAU>  
A:Cross-references: EMBL:Z21972; NID:g288298; PIDN:CAA79985.1; PID:g288300  
A:Experimentally: *Bacillus* cytochrome P450 CYP106; cytochrome P450 homology  
C:Superfamily: *Bacillus* cytochrome P450 CYP106; cytochrome P450 homology  
C:Keywords: chromoprotein; heme; iron; metalloprotein  
C:Binding site: heme iron (Cys) (axial ligand) #status predicted  
C:355/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match	17.0%	Score 371	DB 2	Length 410
Best Local Similarity	30.1%	Pred. No. 9.6e-21		
Matches 109	Conservative 53	Mismatches 180	Indels 20	Gaps
QY	63	WATRGQILREAYEDYRHFSSP-----CPILRAGEAYDPIP-TSMPPROQFRALA	115	
		: : : :   :		:     :   :
DB	44	WNVFVEDYKRVLSDYKKHSSVYKRRTTISVGTDSSEGSVPEKIQTLESDDPRHRRSL	103	
		: : : :   :		:     :   :
QY	116	NOVGMAPVVDKIENNIQIELACSLISLRPGCCNTTEDIAEFPPIRIFMLAGLPREDIP	175	
		: : : :   :		:     :   :
DB	104	AAAFPRSIQWNEPPIQEIADIELIGOMGGTPIDIVASLASPLPIVMADLNGVSPKDL	163	
		: : : :   :		:     :   :
QY	176	HLKYLTDQMTTRDGSMTFAE-----AKEALYDYLPIIEPRROKPGDAISYANGOV	228	
		: : : :   :		:     :   :
DB	164	LEKKAVDILFLPDEKEQEDVKLQVAAKEYQYLLPYVQKRLNPADIIISDLKSEV	223	
		: : : :   :		:     :   :
QY	229	NGRPITSDCAKRMFGILLVGGDITVNVNLSFSM-EFLAKSPREBOELTERPERIPAAE	287	
		: : : :   :		:     :   :
DB	224	DGEMTTDEEVYKTMILLIGAGVETTSILLANSFYSLIDDKKVQGEIHNHNDLVPQAVE	283	
		: : : :   :		:     :   :
QY	268	LIR-RESLVADGRILTSDEYEHGVOLKKGDQILFLPQMSIGDERKNACPMHVSF-RQV	345	
		: : : :   :		:     :   :
DB	284	MIRFEPNLIKIDRYKAEENDLGVELKEGDSVYVWMSAANDMEEMFEDPFTINIRPNK	343	
		: : : :   :		:     :   :
QY	346	SHHTTGHSHLCLGQHLARRELIVLKEMLTPIRPFSTIAPGAQIOHK--SGIVSGVAL	402	
		: : : :   :		:     :   :
DB	344	KHLTFNGPHFLCGAPLARLEAKIKLTAFLKKFKHIEAVPSQLEENLTDSTAGTQTSL	403	
		: : : :   :		:     :   :
QY	403	PL 404		
DB	404	PL 405		
RESULT	4			
S51594				
cytochrome P450 mycG - Micromonospora griseorubida				



```
Query Match      16.3%; Score 355; DB 1; Length 396;
Best Local Similarity 30.1%; Pred. No. 1.5e-19;
Matches 99; Conservative 54; Mismatches 158; Indels 18; Gaps 7;

QY 71 IREAYEDYRHFSSECPFIPREAGEAYDFTPSMDPQORFALANOVGMVVDKLENR 130
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 56 VKRVGDKELFSS-C-MPOQTSSIGNSI-INMDPKRTKIRSVVNAKFPVVKQADEPR 111

QY 131 IOELACSLIESLRQGCNTEDYAEFPPIRITMMLAGLPEEDIPHLKYLTDMT----- 165
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 112 IOETBELLOFGGRSEFDVHDFSPPIVISELLGVPSAHMEQPKASDILLVSTPKD 171

QY 186 -RPGDSMTFAE---AKEALYDYLPIIEEOROKPGTDAISIVANGOVNRPITSDAKR 240
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 172 KSEAEKAFIEERDKCEBELAAFFAGITIEKRKNKPREDDIISIVAEETSEKLSGEEILP 231

QY 241 MFGILLVGLDITVYNFLSFMELFANSPEHROELIERPERIPACEELLR-RESLVADGR 299
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 232 FCTLLVAGNETTNTLISNAMYSLTEPPGYEELRSHPELMPQAVEBALRFRAPAPVLR 291

QY 300 ILTSDFEFGVOLKKGQILLPOMLSGUDERKNACPMHVPSQKVSHTTFGSHCLG 359
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 292 IAKRDEIGHLKEGMVLAFAVASANRDEAKFDRPMFDIRRHPNPIAFGHGHECLG 351

QY 360 QHLARREIIVTLKEMLRIPDF---SIAP 385
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 352 APLARLEANTALISLISAFPHMECVSTTP 380

RESULT
7
cytochrome P450 RV2266 - Mycobacterium tuberculosis (strain H37RV)
N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Mycobacterium tuberculosis
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: F70729
R:Coyle, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; M0ID:98295987
A:Accession: F70729
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-428 <COL>
A:Cross-references: GB:271163; GB:AL123456; NID:g3261610; PIDN:CAB00969.1; PID:e255175;
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV2266
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C:Keywords: oxidoreductase
F:264-401/Domain: cytochrome P450 homology <CYP>

Query Match      15.9%; Score 346.5; DB 1; Length 428;
Best Local Similarity 26.5%; Pred. No. 7.5e-19;
Matches 118; Conservative 72; Mismatches 187; Indels 69; Gaps 17;

QY 5 IOSNANLAPLPVPHVLVDFDMYNPSNLASGVCEAMVLOE-----SNVPDLV 54
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7 IATRVGTF-PRVVP---IADIEL-----GSLDFMALDDVRCGARTTLRREAPISF 54

QY 55 WTR-----CNGHWIATRGQILREAYEDY---RH---FSS-----ECPTIPREA 92
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 55 WPIELPGEVAGNGHWATRK-----YDDVFASRHPDIFSSYPNTITINDQPELAEYF 107

QY 93 GEAYDITPSPMPPEORQORALANOVGMVVDKLENRIOELACSLIESL---RPGQCN 149
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 108 GSM-----IVLDDPRHQRLRSIVSRAFTPRVARIEAIVRAHRLVSSMIANNPPROAD 162
```

```
QY 150 FTEDYAEPPRIETIMLAGLPEEDIPHLKYLTDMT---RPGDSMTFAEAKALYD---Y 203
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 163 LVSELACPLPLQITCDMKGPKADHRIETWTVILGFGPDLATPDEPMOVASIDIGAY 222

QY 204 LPIIEOROKPGTDAISIVANGOVNRPITSDAKRMFGLLVGLDITVYNFLSFMER 263
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 223 ATALAEDRRVNHDDLTSSLVEAEVDGERLSREISFILLVAGNETTNRNATIGVLA 282

QY 264 LAKSPHEROELIERPERI-PAACEELLRRS-LVAGRILTSYEFHGVOLKKGDDIILP 321
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 283 LSRPPEORQWMSDFGLAPTAVEEIVRMASPVYMRRLTODIELRGTKMAAGDKVSLM 342

QY 322 QMLSGDERKNACPMHVDSFROKVSHTTF-GHGSHCLGOHLARREIIVTLKEMLRIPD 380
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 343 YCSANDESKEADPMTFDLARNPNPHLGCGGAHFCCLANLARREIRAFDELRRQMD 402

QY 381 FSIAPGAQIOHKSIVSGVQALPLW 406
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 403 V-VATEEPARLLSQFIHGKITPLVTW 427

RESULT
8
cytochrome P450 - Streptomyces hygroscopicus
C:Species: Streptomyces hygroscopicus
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 31-Jan-2000
C:Accession: T30231
R:Apaticio, J.F.; Molnar, I.; Schwecke, T.; Kenig, A.; Haydock, S.F.; Khaw, L.E.; St Gene 169, 9-16, 1996
A:Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces
A:Reference number: Z20782
A:Accession: T30231
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-404 <APA>
A:Cross-references: EMBL:X86780; NID:g987088; PID:g987105; PIDN:CAA60465.1
C:Genetics:
A:Gene: rapN
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology

Query Match      15.8%; Score 344.5; DB 2; Length 404;
Best Local Similarity 28.6%; Pred. No. 9.8e-19;
Matches 102; Conservative 63; Mismatches 177; Indels 15; Gaps 9;

QY 63 WIATRGQILREAYEDYRHFSSECPF---IPREAGEAYDF-IPPSMDPQORFALANOV 118
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 50 WLAASWEDVAKFVDPDR-FSRSATLIGKDVPRVLPATIQDQPVIMLMDPPEHTRRLRVATKA 108

QY 119 VGMVVDKLENRIOELACSLIESLRPOG-OCNTEDYAEFPPIRITMMLAGLPEEDIPHL 177
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 109 LNSRRMEALRPRTQVEYADLLIDMKLAKGAPADIMEDFALPLPTIMICELLGVPIDEQTYF 168

QY 178 KYLTDMTRPDGSMT---FAEAKALYDYLPIIEEOROKPGTDAISIVANGOVNRPIT 234
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 169 RTWSDQML-SNGAYSQEVYMAAGQSLYLSIELIARRRQDPTNDLIGSLVRADKDRIS 227

QY 235 SDEAKRMFGLLVGLDITVYNFLSFMELFANSPEHROELIERPERIPACEELLRRSL 294
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 228 ERELWGAFTVTLTLAGETANMIGNSVYTLTHPEKTLAKRLDLSIPRAVELLRIIPI 287

QY 295 VADG---RLTSDFEFGVOLKKGQIILLPOMLSGIDEKKNACPMHVDSFROKVSHTTFG 351
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 288 AKQASVNRMAVEDVELSGITVAKGEAVAIQTSANTDPVYOHPEIDYHRTSNPMSIG 347

QY 352 HGSHCLGOHLARREIIVTLKEMLRIPDFSIAPGAQIOHKSIVSGVQALPLW 406
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 348 HGAHHCMAQIVRVENQGTALGSLISRIPLARVAPVPRKIFLGRILVPSLEALPLTW 404

RESULT
9
B35401
```



QY WIATGQGLIREVEYEDYRHRSSE---CPPIPRACGAYOPIP-----TSMDDPPOPOERLAA 115  
63  
Db MNVFOYEHVKOVLNVYDFSSDGGQRTTIFVGDNSKKKSTSPITNLNTINDPDRHKARSL 104  
45  
QY NOVVGMPVYDKLKNFIQELACSLIESLRPOGOCNFTEDYAEBFPIRITMLLAGPEEDIP 175  
116  
Db AAATPRSLKMWEPRIKIAADLVEAIKONSTINIVDDLSPSPSLIADLFGVGVKDRY 164  
105  
QY HLKYLTDQMTPRDGSMTFAAKE-----ALYDLPIIEORROKPGTDAISYANOV 228  
176  
Db QFKKVVLDLPQYDPERLEETIQEKGKRGAGEFYGLYIVIEKRNLSDDDIISDLIQAEV 224  
165  
QY NGRPTTSDAKRMFCLLVGLDVIYVNF-L-SFSMEPLAKSPBHROELIERBRIPACEE 287  
229  
Db DGEFTTDEEIHATMLLIGAGVETTSIAIANNFYSLDLDKSLYSELNNNELAPKAVEE 284  
225  
QY LLR-RESVLVADGRILTSYEFHGVOLTKGGDILLQMLSGDERKNACPMHVDFSR-QKV 345  
288  
Db MLRYRFHLSRRDRIVKQONELLGVKLKGDGVYIAAMSAACNMDETFFEMPFSDVITHRPINK 344  
285  
QY SHTTTGHSHLCLGQHLARRELIYVLKEMLTR-----IPDFSIAIP 385  
346  
Db KHLTFGNGPHFCGLPIARLEMKIILEAFLEAFSHIEPFEDELEIP 390  
345

RESULT 12  
E69611  
cytochrome P450 cypA - *Bacillus subtilis*  
N:Contains: oxidoreductase (EC 1.-.-.-)  
C:Species: *Bacillus subtilis*  
C:Date: 10-Sep-1999 #sequence.revision 10-Sep-1999 #text.change 18-Feb-2000  
C:Accession: E69611; T44774  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertet  
C.; Bron, S.; Broillett, S.; Bruschi, C.V.; Caldwell, V.; Capuano, V.; Carter, N.M.; Ch  
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Gall  
lech, J.; Harwood, C.R.; Henauf, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
Koether, P.; Konligstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Ladi  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue  
Y. M.; Ogawa, K.; Ogizawa, A.; Oudega, B.; Park, S.H.; Parvo, V.; Pohl, T.M.; Port  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon  
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot  
akenchul, M.; Yamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama  
T.; Winters, P.; Wiput, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida  
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, K.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
A:Reference number: A69580; MUID:98044033  
A:Accession: E69611  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA

A:Cross-references: GB:599117; NID:92634966; PIDN:CAB14615.1; PID:el183903;  
A:Experimental source: strain 168  
J.Belitsky, B. R.; Gustafsson, M.C.U.; Sonnenshein, A.L.; von Wachenfeldt, C.  
J. Bacteriol. 179, 5448-5457, 1997  
A:Title: An lrp-like gene of *Bacillus subtilis* involved in branched-chain amino acid tra  
A:Reference number: 22837  
A:Accession: T44774  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1410 <BEL>  
A:Cross-references: EMBL:Y11043; PIDN:CAA71937.1  
A:Experimental source: strain 1A1  
C:Genetics:  
A:Gene: cypA  
C:Superfamily: *Bacillus* cytochrome P450 CYP106; cytochrome P450 homology  
C:Keywords: oxidoreductase  
C:Keywords: 245-381/Domain: cytochrome P450 homology <CYP>  
C:245-381/Domain: cytochrome P450 homology <CYP>

Query Match	15.38;	Score 334;	DB 1;	Length 410;
Best Local Similarity	27.28;	Pred. NO. 6.3e-18;		

[illegible]

RESULT 13  
B42606  
cytochrome P450 CYP11B1 - Saccharopolyspora erythraea  
N:Contains: oxidoreductase (EC 1.-.-.-)  
C:Species: Saccharopolyspora erythraea  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 03-Mar-2000  
C:Accession: B42606  
R:Andersen, J.F., Hutchinson, C.R.  
J. Bacteriol. 174, 725-735, 1992  
A:Title: Characterization of Saccharopolyspora erythraea cytochrome P-450 genes and  
A:Reference number: A42606; MUID:92121109  
A:Accession: B42606  
A:Status: preliminary: not compared with conceptual translation  
A:Molecule type: DNA; protein  
A:Residues: 1-405 <AND>  
A:Cross-References: GB:883110, NID:g152682, PIDN:AAA26483.1; PID:g152684  
A:Experimental source: NRR12338  
A:Note: sequence extracted from NCBI backbone (NCBIP:77484)  
C:Genetics:  
A:Gene: CYP107B1  
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology  
C:Keywords: chromoprotein; heme; Iron; metalloprotein; oxidoreductase  
C:328-3/4/domain: cytochrome P450 homology <CYP>  
C:352/Blinding site: heme iron (Cys) (axial ligand) #status predicted

Query Match	15.3%	Score 333.5	DB 1	Length 405
Best Local Similarity	25.6%	Pred. No. 6.7e-18		
Matches 107	Conservative 74	Mismatches 104	Indels 53	Gaps 10

  

QY	18	VPEHLVFD---	FDMYPSNLSAGVQEMAAVLQESNPDLVWTRCNG-GHNIATRGQLIR	72
		:     :	:     :	
Db	6	VPDLIAFDAQOQRHN-----	RYARMREBPVQRI--RTVNGGLDAWLIRREYDK	53
		:     :	:     :	
QY	73	EAYEDYRHFSECFPIRREAGEATDPIPTSM-----	DPEQROQFR	112
		:     :	:	
Db	54	QALLD-----PRIADFGERTQOILIEKRLADERRRFGSPDLGPHMLNTDPDPDHRRLR		105
		:     :	:	
QY	113	ALANQVGMVYVDKLENNRIQACLSLIESLRPOGOCNFTEDYAEFPRIPIRLMAGLPEE		172
		:     :     :     :     :     :	:     :	
Db	106	KLVYKAFARAREVEGLRIPEQITDDLLDRLAGRSVDDIDEFARPLPTIVSIELMGVSDS		165
		:     :     :     :     :     :	:     :	
QY	173	DIPHLKYLITDMDTPDGSMTFAEAKE--ALDYDILPIIEQRORRPGDTAISIVANGVNG		230
		:     :     :     :     :     :	:     :	
Db	166	RRDFRSTNTVLV--DGSQPEAQAAQASVAMVEYLTLELLAKRTSGDGLLTALLAEAVDG		223
		:     :     :     :     :     :	:     :	







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 13:04:53 ; Search time 27.02 Seconds

(Without alignments)  
474.815 Million cell updates/sec

Title: US-09-246-451-13

Perfect score: 2177  
Sequence: 1 TTERIQSNANLAPLPHPVPE.....IVSGVOALPLWDPATRAV 414

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 85661 seqs, 30989116 residues

Total number of hits satisfying chosen parameters: 85661

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_38.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2165	99.4	414	1 CPXA_PSEPU	P00183 pseudomonas
2	591.5	27.2	414	1 Y4VG_RHISN	O53215 rhizobium s
3	376	17.3	393	1 YH85_MYCTU	O53936 mycobacteri
4	371	17.0	410	1 CPXM_BACME	O06069 bacillus me
5	359	16.5	398	1 YZ1B_MYCTU	O53563 mycobacteri
6	355	16.3	396	1 YZ1B_MYCTU	O34374 bacillus su
7	350.5	16.1	405	1 CPXM_BACSU	P27632 bacillus su
8	346.5	15.9	428	1 YM66_MYCTU	O50696 mycobacteri
9	340	15.6	402	1 CPXF_STRGO	P18327 streptomyc
10	334.5	15.4	410	1 CPXL_PSEPU	P14762 bacillus me
11	334	15.3	410	1 CPXL_PSEPU	O08469 bacillus su
12	333.5	15.3	405	1 CPXM_SACER	P33271 saccharopol
13	328	15.1	405	1 CPXE_STRGO	P18326 streptomyc
14	323	14.8	428	1 CPXL_PSEPU	P33006 pseudomonas
15	318.5	14.6	405	1 YC56_MYCTU	O11062 mycobacteri
16	317.5	14.6	399	1 FAS1_RHOFA	P63733 rhodococcus
17	314	14.4	412	1 CPXH_STRGR	P53554 bacillus su
18	305.5	14.0	395	1 BIOT_BACSU	P26911 streptomyc
19	302	13.9	458	1 Y180_MYCTU	O08464 mycobacteri
20	301	13.8	402	1 NOR_FUSOX	P33295 fusarium ox
21	294.5	13.5	408	1 NOR_CYLTO	O12599 cylindrocac
22	293	13.5	381	1 CPXG_STRSO	P33266 streptomyc
23	291.5	13.4	433	1 YZ45_MYCTU	P18856 mycobacteri
24	286.5	13.2	403	1 NOR_CYLTO	O00616 cylindrocac
25	286.5	13.2	404	1 CPXJ_SACER	O00441 saccharopol
26	276.5	12.7	397	1 CPXO_SACER	P48655 saccharopol
27	273	12.5	414	1 Y778_MYCTU	P77903 mycobacteri
28	263	12.1	400	1 CPXP_RHISN	P55544 rhizobium s
29	251	11.5	401	1 CPXP_RHISN	O59203 bradyrhizob
30	237	10.9	402	1 Y766_MYCTU	P77992 mycobacteri
31	235	10.8	396	1 YM76_MYCTU	O59571 mycobacteri
32	225	10.3	436	1 THCB_RHOER	P43492 rhodococcus
33	221.5	10.2	400	1 YV21_MYCTU	O08362 mycobacteri

34	218	10.0	422	1 CPXC_AGR76	P24466 agrobacteri
35	216	9.9	429	1 CPXR_BRAJA	O59204 bradyrhizob
36	204.5	9.4	489	1 YM68_MYCTU	O59572 mycobacteri
37	196	9.0	405	1 CPXP_BACSU	O34926 bacillus su
38	190	8.7	467	1 CPXR_RHISN	P55543 rhizobium s
39	178	8.2	368	1 CPXM_SULSO	O55080 sulfolobus
40	163	7.5	1048	1 CPXB_BACME	P14779 bacillus me
41	155.5	7.1	503	1 CP30_SHEEP	O29496 ovis aries
42	155.5	7.1	531	1 CP27_HUMAN	O02318 h cytochrom
43	152	7.0	407	1 CPXD_AGR76	P24467 agrobacteri
44	151	6.9	492	1 CP33_PIG	O02390 sus scrofa
45	146.5	6.7	489	1 CP33_RABIT	P00182 oryctolagus

## ALIGNMENTS

RESULT	1	STANDARD	PRT	414 AA.
CPXA_PSEPU				
ID	CPXA_PSEPU			
AC	P00183:			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	13-AUG-1987 (Rel. 05, Last sequence update)			
DT	15-FEB-2000 (Rel. 39, Last annotation update)			
DE	CYTCHROME P450-CAM (EC 1.14.15.1) (CAMPOR 5-MONOOXYGENASE).			
GN	CAMC OR CYP101.			
OS	Pseudomonas putida.			
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;			
OC	Pseudomonas.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE: 86111751.			
RA	Unger B.P., Gunsalus I.C., Sliagar S.G.;			
RT	"Nucleotide sequence of the Pseudomonas putida cytochrome P-450cam			
RT	gene and its expression in Escherichia coli.";			
RL	J. Biol. Chem. 261:1158-1163(1986).			
RN	[2]			
RP	SEQUENCE OF 385-414 FROM N.A.			
RC	STRAIN=ATCC 17453.			
RX	MEDLINE: 90130389.			
RA	Koga H., Yamaguchi E., Matsunaga K., Aramaki H., Horinuchi T.;			
RT	"Cloning and nucleotide sequences of NADH-putidaredoxin reductase			
RT	gene (camA) and putidaredoxin gene (camB) involved in cytochrome			
RT	P-450cam hydroxylase of Pseudomonas putida.";			
RL	J. Biochem. 106:831-836(1989).			
RN	[3]			
RP	SEQUENCE.			
RX	MEDLINE: 83030788.			
RA	Haniu M., Ames L.G., Yasunobu K.T., Shastri B.A., Gunsalus I.C.;			
RT	"Amino acid sequence of the Pseudomonas putida cytochrome P-450. II.			
RT	Cyanogen bromide peptides, acid cleavage peptides, and the complete			
RT	sequence.";			
RL	J. Biol. Chem. 257:12664-12671(1982).			
RN	[4]			
RP	X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).			
RX	MEDLINE: 86143817.			
RA	Finzel B.C., Weber P.C., Hardman K.D., Salame F.R.;			
RT	"Structure of ferricytochrome c' from Rhodospirillum rubrum at			
RT	1.67-A resolution.";			
RL	J. Mol. Biol. 186:627-643(1985).			
RN	[5]			
RP	X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).			
RX	MEDLINE: 86059514.			
RA	Poulos T.L., Finzel B.C., Gunsalus I.C., Wagner G.C., Kraut J.;			
RT	"The 2.6-A crystal structure of Pseudomonas putida cytochrome P-450.";			
RT	J. Biol. Chem. 260:16122-16130(1985).			
RN	[6]			
RP	X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).			
RX	MEDLINE: 98019009.			
RA	Schlichting I., Jung C., Schultze H.;			
RT	"Crystal structure of cytochrome P-450cam complexed with the (1S)-			
RT	camphor enantiomer.";			
RL	FEBS Lett. 415:253-257(1997).			

RN [7]  
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).  
RX MEDLINE; 98313255.  
RA Vidakovic M., Sligar S.G., Li H., Poulos T.L.;  
RT "Understanding the role of the essential Asp251 in cytochrome p450cam  
RT using site-directed mutagenesis, crystallography, and kinetic solvent  
RT isotope effect.";   
RL Biochemistry 37:9211-9219(1998).  
RN [8]  
RP STRUCTURE BY NMR.  
RX MEDLINE; 97459726.  
RA Moura C., Bondon A., Simmoneaux G., Jung C.;  
RT "1H-NMR study of diamagnetic cytochrome P450cam: assignment of heme  
RT resonances and substrate dependance of one cysteinate beta proton.";   
RL FEBS Lett. 414:203-208(1997).  
CC -1- FUNCTION: INVOLVED IN A CAMPHOR OXIDATION SYSTEM.  
CC -1- CATALYTIC ACTIVITY: CAMPHOR + PUTIDAREDOXIN + O(2) = 5-EXO-  
CC HYDROXYCAMPHOR + OXIDIZED PUTIDAREDOXIN + H(2)O.  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M12546; AAA25760.1; -;  
CC DR EMBL; D00528; BAA00412.1; -;  
CC DR PIR; A25660; OAPSCP.  
CC DR PDB; 2COP; 15-APR-91.  
CC DR PDB; 3COP; 15-APR-91.  
CC DR PDB; 4COP; 15-JUL-91.  
CC DR PDB; 5COP; 15-JUL-91.  
CC DR PDB; 6COP; 15-JUL-91.  
CC DR PDB; 7COP; 15-JUL-91.  
CC DR PDB; 8COP; 15-JUL-91.  
CC DR PDB; 1CP4; 15-JAN-93.  
CC DR PDB; 2CP4; 15-JAN-93.  
CC DR PDB; 3CP4; 15-JAN-93.  
CC DR PDB; 4CP4; 15-JAN-93.  
CC DR PDB; 5CP4; 15-SEP-98.  
CC DR PDB; 6CP4; 16-SEP-98.  
CC DR PDB; 1NOO; 08-MAR-96.  
CC DR PDB; 1PHA; 31-OCT-93.  
CC DR PDB; 1PHB; 31-OCT-93.  
CC DR PDB; 1PHC; 31-OCT-93.  
CC DR PDB; 1PHD; 31-OCT-93.  
CC DR PDB; 1PHE; 31-OCT-93.  
CC DR PDB; 1PHF; 31-OCT-93.  
CC DR PDB; 1PHG; 31-OCT-93.  
CC DR PDB; 1AKD; 19-NOV-97.  
CC DR PFAM; PF00067; P450; 1.  
CC DR PROSITE; PS00066; CYTOCHROME\_P450; 1.  
CC KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;  
CC KM 3D-structure.  
CC FT INIT\_MET 0  
CC FT BINDING 357 0  
CC FT CONFLICT 55 56 HEME.  
CC FT CONFLICT 276 276 E-> Q (IN REF. 3).  
CC FT CONFLICT 361 361 H-> S (IN REF. 3).  
CC FT CONFLICT 407 407 D-> N (IN REF. 3).  
CC FT TURN 16 17  
CC FT HELIX 20 22  
CC FT STRAND 23 23  
CC FT TURN 28 29  
CC FT TURN 32 33  
CC FT HELIX 34 36  
CC FT HELIX 38 46  
CC FT TURN 48 49  
CC FT STRAND 53 56  
CC FT HELIX 58 60

FT	STRAND	62	65
FT	HELIX	68	76
FT	TURN	78	80
FT	STRAND	81	82
FT	TURN	83	84
FT	HELIX	90	95
FT	TURN	99	102
FT	TURN	105	107
FT	HELIX	108	119
FT	HELIX	121	145
FT	TURN	146	146
FT	STRAND	147	149
FT	HELIX	150	153
FT	TURN	154	156
FT	HELIX	157	167
FT	TURN	168	168
FT	HELIX	171	185
FT	HELIX	193	213
FT	HELIX	219	224
FT	TURN	225	225
FT	STRAND	227	228
FT	TURN	229	230
FT	STRAND	231	232
FT	STRAND	235	250
FT	HELIX	251	252
FT	TURN	253	266
FT	HELIX	268	276
FT	HELIX	278	291
FT	TURN	292	292
FT	STRAND	295	295
FT	STRAND	297	301
FT	STRAND	305	307
FT	TURN	308	309
FT	STRAND	310	312
FT	TURN	314	315
FT	STRAND	317	320
FT	HELIX	322	324
FT	TURN	325	327
FT	TURN	329	331
FT	TURN	335	336
FT	TURN	340	341
FT	TURN	349	350
FT	HELIX	353	355
FT	TURN	358	359
FT	HELIX	360	377
FT	STRAND	382	383
FT	TURN	385	386
FT	STRAND	391	392
FT	STRAND	396	396
FT	STRAND	398	399
FT	STRAND	403	405
FT	HELIX	408	410
SO	SEQUENCE	414 AA; 46538 MW; BCA610293A9D6207 CRC64;	

Query Match 99.4%; Score 2165; DB 1; Length 414;  
Best Local Similarity 99.5%; Pred. No. 4.8e-156;  
Matches 412; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY	1	TTETTSNANLAPLPHVBEHLVDFDMDNPSNLSAGVDEANAVLQESNVPDLVWTRCG	60
DB	1	TTETISNANLAPLPHVBEHLVDFDMDNPSNLSAGVDEANAVLQESNVPDLVWTRCG	60
QY	61	GHVIATRGOLIREAYDYHNFSECFIPREAGEAYDFIPTSMDDPEORQFRALANQVVG	120
DB	61	GHVIATRGOLIREAYDYHNFSECFIPREAGEAYDFIPTSMDDPEORQFRALANQVVG	120
QY	121	MPVVDKLENRIQELACSLIESLRPQGCFTEEDYAPFPFIRIFMLLAGLPEEDIPLKYL	180
DB	121	MPVVDKLENRIQELACSLIESLRPQGCFTEEDYAPFPFIRIFMLLAGLPEEDIPLKYL	180
QY	181	TDQMTFPGSGMTFAEAKALYDYLPIITQOROKPGTDAISIVANGQVNGRPITTSDEAKR	240

Db	181	TDQMPRPDSMTFAEAKFALDYLLPIETQRQRKGTDAISVANGQVNGRITSDENAKR	240
Qy	241	MGGLLVAGGLDTVNVNPLFSFMEFLAKSPBHRQELIERPERIPACEBELLRRPSLVADGRI	300
Db	241	MGGLLVAGGLDTVNVNPLFSFMEFLAKSPBHRQELIERPERIPACEBELLRRPSLVADGRI	300
Qy	301	LTSDFEFGVGLKKDDQDILLPOMLSGLDERKACPMHDFSRQXSHTEFGGSHLCGQ	360
Db	301	LTSDFEFGVGLKKDDQDILLPOMLSGLDERKACPMHDFSRQXSHTEFGGSHLCGQ	360
Qy	361	HLARREIIVTLKEMLTIRIPDPSIABGAOIOHKSGIVSGVQALPLVMDPATTTKAV	414
Db	361	HLARREIIVTLKEMLTIRIPDPSIABGAOIOHKSGIVSGVQALPLVMDPATTTKAV	414
RESULT	2		
YAVG_RHISN	YAVG_RHISN	STANDARD:	PRT: 414 AA.
ID	053215:		
AC	01-NOV-1997	(Rel. 35, Last sequence update)	
DT	01-NOV-1997	(Rel. 35, Last sequence update)	
DT	15-DEC-1998	(Rel. 37, Last annotation update)	
DE	PROBABLE CYTOCHROME P450 YAVG (EC 1.14.14.-).		
GN	YAVG.		
OS	Rhizobium sp. (strain NGR234).		
OG	Plasmid sym PNGR234a.		
OC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;		
CC	Rhizobiaceae; Rhizodium.		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE: 97305956.		
RA	Freiberg C.A., Fellay R., Balroch A., Broughton W.J., Rosenthal A.,		
RA	Perret X.		
RT	"Molecular basis of symbiosis between Rhizobium and legumes."		
RL	Nature 387:394-401(1997).		
RN	(2)		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE: 96389014.		
RA	Freiberg C., Perret X., Broughton W.J., Rosenthal A.		
RT	"Sequencing the 500-kb GC-rich symbiotic replicon of Rhizobium sp.		
RT	NGR234 using dye terminators and a thermostable 'sequenase': a		
RT	beginning."		
RL	Genome Res. 6:590-600(1996).		
CC	-I- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE		
CC	MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED		
CC	COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.		
CC	-I- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation		
CC	the European Bioinformatics Institute. There are no restrictions on		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see <a href="http://www.isb-slb.ch/announce/">http://www.isb-slb.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@isb-slb.ch">license@isb-slb.ch</a> ).		
CC	-----		
DR	EMBL: Z68203; CA92422.1; -		
DR	EMBL: AE000101; AAB91893.1; -		
DR	HSSP: P00183; 3CP4.		
DR	PFAM: PF00067; p450_2.		
KW	PROSITE: PS00086; CYTOCHROME P450; FALSE NEG.		
KW	Hypothetical protein: Oxidoreductase; Monooxygenase;		
KW	Electron transport; Heme; Plasmid.		
FT	BINDING 364 364 HEME (BY SIMILARITY).		
SEQUENCE	414 AA: 45810 MW: 888598E99315BB84 CR664;		

	Query Match	27.2%	Score 591.5;	DB 1;	Length 414;
	Best Local Similarity	34.1%;	Pred. No. 1.8e-37;		
	Matches 142; Conservative	65;	Mismatches 171;	Indels 39;	Gaps 9;
QY	13 PLPHVPELVDEDFMKNPSLSA---GYQEAMAVLQESNVDDLYW----TRCNGSHMTA 65				
	:               :	:	:	:	:           :

```

Db 13 PFDHVPALVYRHHSLFSPGMAPTPNDDPHAAVACHDDGPRTFFSPSNTROGRGTWY 72
Qy 66 TRGQLIREAYEDYRHHFSSECFPIRENGEAYDFITPSMDPEEORQFALANOVVGMPEVD 1355
Db 73 TRARQORFVLTEDTETESSHRSIFASALGEMHPVPIELDPHAGVFALLNPIFSSRYL 1322
Qy 126 KLENRIQELACSLIESL-RPGGOCNFIEDYAEPPRIPIFMLAGLP-----EDT 174
Db 133 ALEPTIHARACALIDCLAKETSCDYMKDFALPTFSVLSFLIGLSORREVLGWVSDL 1922
Qy 175 PLKLVLTQGMTRPGOSMFAEAKELAYDLPIIEO-----RROKPGDAISYANGVNG 230
Db 193 LH-----GN-----AEKRRAARSVAATIDEMANMRKSPAYDEMTFVQAITEG 2377
Qy 231 RPIITSEAKRMFGLLVGGDLTVVNFLSFSMEFLAKSPENRQELTERPERIPACCEILR 290
Db 238 RSLTBEERYRGIVFLVAGDITVAALIGFDMAVLYARNKQHOLLNRNPARIGLAABELLR 297
Qy 291 RFSLVADGRILTSDYEFNGVOLKKGDQILLPOMLSGLDERKNACPMHVDFSROKVSHTTF 3506
Db 298 AVSTVQOILRVATKDLIEFGCVPIREGDYVSCPAMLANRNPSEFCRPTNIDLARODNHTAF 3578
Qy 351 GHGSLGCGHMLAREIIVTLKEMTLTRPDPSIAPG-AQIOHKSIGSIYGOALPLWY 406
Db 358 GIGPLHCGAHARREIYIGLRERWAKRIPARIEGRAPTH-GGHVFGISNILLW 413

```

RESULT	3
YH85_MCMCU	
ID	YH85_MACTU
STANDARD:	PRT: 393 AA.
AC	053936;
DT	15-FEB-2000 (Rel. 39, Created)
DT	15-FEB-2000 (Rel. 39, Last sequence update)
DT	15-FEB-2000 (Rel. 39, Last annotation update)
DE	PUPATIVE CYTOCHROME P450 RV1785C (EC 1.14.-.-).
GN	RV1785C OR MTV049.07C.
OS	Mycobacterium tuberculosis.
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-H37RV.
RX	MEDLINE: 98295987.
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA	Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA	Baddock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA	Holmesby T., Jagels K., Krogh A., Mclean J., Moule S., Murphy L.,
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA	Rutter S., Seeger K., Skellern S., Squares S., Squares R., Sulston J.E.,
RA	Taylor K., Whitehead S., Barrall B.G.;
RT	"Deciphering the biology of Mycobacterium tuberculosis from the
RT	complete genome sequence.";
RL	Nature 393:537-544(1998).
CC	- 1. SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL: AL022021, CA11707.1; -.
DR	TUBERCULIST: RV1785C; -.
DR	PFAM: PF00067; p450; 2.
DR	PROSITE: PS00086; CYTOCHROME_P450; 1.
KW	Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.
PT	BINDING 342 342 HEME (BY SIMILARITY).
SO	SEQUENCE 393 AA: 43541 MW: 860DD1504471FAB8 CCK64;

Query Match	17.3%: Score 376; DB 1; Length 393;
Query Local Similarity	28.7%: Pred. No. 3e-21;
Matches 110; Conservative 70; Mismatches 183; Indels 20; Gaps	
QY 31	PSNLASGVQEAAMAVLQESNVSDLVWTRCGNGHWIATRGQILREAYEDYRHFSSCEPIPR 90
DB 23	PMADVGRV--GMKTLRDAG--PVVF--MNGWYTL--TRREDVIALRNPKEVSSKALQP- 74
QY 91	EAGAYEDIPFISMPREDPQCFRALANOVGMPVVDKLENRQELACSLISLRQGCQNF 150
DB 75	-PGNPLPVPLAFDPPEHTRRILOPYFSPALSKALPSSLRRTVAMITAIAGRGCEA 133
QY 151	TEDVAEPPIPIFIMLAGLPREDIPHLKYTLDDQ---TRPDGSMTEAEKAEALYDLIP 206
DB 134	MADLANLFPOFLVLYGLPLIEDRDRLGKMDAVIANSDRPHPEADVAARLELTETA 195
QY 207	TEORROKPGTDAISIVANGOVNRPITSDCAKRMFGILLVGIDTVVNFPSFMEELAK 266
DB 194	MVAERRRNPBGPDVLSQV---QIGDPDISEIVLDSHLLIAGLDYTAAVGFSLEELAR 250
QY 267	SPENHOLEIPEPERIPAAACEELLR-RFSLVADGRILTSDVEFGVQLKKGDQILLPMLS 325
DB 251	RPOLRAMLRDNPKOIRVFIEIIVLEPSAPVAPRVPTTEPVVGMTLPAGSPVRLCAAAV 310
QY 326	GLDERKKNACPMHNPVDSRQKVSHTFEGSHLCLGONHARREIYTLKEMLRIPDESIAF 385
DB 311	NRDSDAMSTDELMDVGIVRHMGVGGGPHRCLSHLARLETLVLGEMNLQIPDEELAF 370
QY 386	--GAQIOHKSGIVSGVALPLVW 406
DB 371	DYAPEIRREPSKSPA-LKRLPLRW 392

  

RESULT 4	
CPXM_BACME	STANDARD; PRT; 410 AA.
ID	CPXM_BACME
AC	006069;
DT	01-FEB-1995 (Rel. 31, Created)
DT	01-FEB-1995 (Rel. 31, Last sequence update)
DT	15-FEB-2000 (Rel. 39, Last annotation update)
DE	CYCLOCHROME P450(ME) (EC 1.14.99.-) (STEROID 15-BETA-HYDROXYLASE)
DE	(STEROID 15-BETA-MONOXYGENASE).
GN	CYP106A2.
OS	Bacillus megaterium.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;
OC	Bacillus/Staphylococcus group; Bacillus.
RP	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-ATCC 13368;
RA	MEDLINE; 94049677.
RA	Rauschenbach R., Isernhagen M., Noeske-Jungblut C., Boidol W.,
RT	Siewert G.;
RT	*Cloning sequencing and expression of the gene for cytochrome
RT	P450meg, the steroid-15 beta-monoxygenase from Bacillus megaterium
RT	ATCC 13368.";
RL	Mol. Gen. Genet. 241:170-176(1993).
RL	[2]
RP	CHARACTERIZATION.
RP	STRAIN-ATCC 13368;
RC	MEDLINE; 79194051.
RA	Berg A., Ingelman-Sundberg M., Gustafsson M.;
RA	*Purification and characterization of cytochrome P-450meg.";
RT	J. Biol. Chem. 254:5264-5271(1979).
RL	[3]
RP	CHARACTERIZATION.
RP	STRAIN-ATCC 13368;
RC	MEDLINE; 82091079.
RA	Berg A., Rafter J.J.;
RA	*Studies on the substrate specificity and inducibility of cytochrome
RT	P-450meg.";
RL	Biochem. J. 196:781-786(1981).
CC	-1- FUNCTION: HAS THE CAPACITY TO HYDROXYLATE CERTAIN STEROIDS IN THE
CC	15-BETA POSITION.

```

CC      -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; Z21972; CAAT9985.1; -.
DR      PIR; S32216; S32216.
DR      PIR; S39924; S39924.
DR      PIR; S39924; S39924.
DR      PIR; S39924; S39924.
DR      PIR; S39924; S39924.
DR      PRINTS; PR00359; BP450.
DR      PROSITE; PS00086; CYTOCHROME_P450.1.
DR      OXIDOREDUCTASE; Monooxygenase; Electron transport; Membrane; Heme.
KW      BINDING 355 355 HEME (BY SIMILARITY).
FT      BINDING 355 355 HEME (BY SIMILARITY).
SO      SEQUENCE 410 AA; 46955 MW; 9PDDDCFE2E0F810B CRC64;

Query Match 17.0%; Score 371; DB 1; Length 410;
Best Local Similarity 30.1%; Pred. No. 7.5e-21;
Matches 109; Conservative 53; Mismatches 180; Indels 20; Gaps 7;

QY      63 WIAIRGQILREAYEDYRHFSSE-----CPPIPRAGEAYDFIP-TSKMPPEQORFALA 115
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      44 WNVKREYEDYKRVLSIDYKHFSSVRRKRTTISVCTDSEEGSVPERKIQTTESDPDHRKRSL 103
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      116 NQVVGMPVVDKLENRIOELACSLIESLRPGOCNTEYDAPRPRIIMLAGLPEEDIP 175
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      104 AAATPNSLQWPERIOLIADELIGOMGTEIDIVASLASLPRIIVADLMGVPSKRL 163
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      176 HLKYLTDQMTPRDPSMTFAE-----AKAALDYLLPIIQROROKPCTDAISYANGOV 228
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      164 LFKWVDLFLPFRBKEQEVYDKLQVAKKEYQYLYIVVQKRLNPRADLIISLASEV 223
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      229 NGRPTISDEAKRMFGLLVGLDVTVNFLSFM-EFLAKSPENROELIERPERIPACEE 287
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      224 DGEHFTDDEVYRTMILLIGAGVETTSHLANSFSLYDDEKVEYQELHENDILVQAVEE 283
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      288 LLR-RESLVADGRILTSYEFHGVQVQKKGDDILLPOMLSGDERKNACPMYVDSR-QKV 345
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      284 MLRRFENLIKLDRIYKENDLIGVELLEGDSVVMMSAANDEEFEDPFTLIHPRNK 343
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      346 SHTEFGHSHLCTOHLARREIYTLKEMLRIPDEFSLAPAOIQHK---SGIVSGVAL 402
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      344 KHLTFGNGPHFCGLAPLARLEAKIALTLFLFKFKHIEAVPFOLEBNLTDSATGOTLTSL 403
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      403 PL 404
      : :
DB      404 PL 405

RESULT 5
YZ18_MYCTU STANDARD; PRT; 398 AA.
ID YZ18_MYCTU STANDARD; PRT; 398 AA.
AC 053563;
DT 15-FEB-2000 (Rel. 39, Created)
DT 15-FEB-2000 (Rel. 39, Last sequence update)
DE 15-FEB-2000 (Rel. 39, Last annotation update)
DE PUTATIVE CYTOCHROME P450 RV3518C (EC 1.14.-.-).
GN RV3518C OR MTVO23.25C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV.
RC MEDLINE; 98295987.
RA Cole S.T., Broesch R., Parkhill J., Garner T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,

```

RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krog A., Moles J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skellern S., Squares S., Squires R., Sulston J.E.,  
RA Taylor K., Whitehead S., Barrett B.G.,  
RT "Deciphering the Biology of Mycobacterium tuberculosis from the  
RT complete genome sequence."  
RL Nature 393:537-544(1998).  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AL022022; CAI17755.1; -.  
CC TUBERCULIST; RV3518C; -.  
CC PFAM: PF00067; P450; 1.  
CC PRINTS: PR00359; BP450.  
CC PROSITE: PS00086; CYTOCHROME\_P450; 1.  
CC DR Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.  
CC KW BINDING 340 340 HEME (BY SIMILARITY).  
CC FT BINDING 340 340 HEME (BY SIMILARITY).  
CC SQ SEQUENCE 398 AA; 44398 MW; BCFP3C23ECB5767F CRC64;

Query Match 16.5%; Score 359; DB 1; Length 398;  
Best Local Similarity 28.2%; Pred. No. 5; 8e-20;  
Matches 107; Conservative 69; Mismatches 169; Indels 34; Gaps 8;

QY 55 WTFCN-----GGHWATGSGOLIREAYEDYRHFS-----ECPTFRAGENAYDPT 101  
DB 26 WMRANQVDFDRNGLAASYQAVIDERQPELFSNAGIRPDQALP-----MMT 76  
QY 102 SMPPEORFALANQVGVVVDKLENRIOELACSLIESLRPOGOCNFTEDYAEPPIR 161  
DB 77 DMDPAILLRKRLVNGFTTKRKVKDKASTALCDTLIDAVGERGEDPFDLAAPLMA 156  
QY 162 IFMLAGL-PEEDIPHLK-----YLTDMTRPGSMTFAEAKALYDILPIIEQRQ 213  
DB 137 VIDMDLGVREQDMFLRMSDDLVTFLSHVSODEFOITM-DAFAAYNDFTRAIARRA 195  
QY 214 KPGTDAISIVANGVNGRPTISDEAKRMFGLLVGLDVTNVPLSFMEFLAKSPEAROE 273  
DB 196 DPTDDLVSVSVSEVDEERISDDELVMETLILIGDETRHTLSGGTEQLLRDDQWL 255  
QY 274 LIERPERIPACEELLRRFSLVAD-GRILTSDEYFHVQQLKGDQIILPQMLSGIDERN 332  
DB 256 LQNDPSLLPAIEMLRWTPVKNMCRVLTADIEFHGTALCAGEKMLLEPSANFEDAVF 315  
QY 333 ACPHNVDFSQKVSHTTGSHGLCLGQHLARREIYTLKEWLTTRIPDFSIAPGAQI--Q 390  
DB 316 CEPEKFPVQGRNPNSHLAFGFGTHCLGNQALRELSLMTERVLRRLPDLRLVADSVPL 375  
QY 391 HKSGIVSGVALPLVMDPA 409  
DB 376 RPANFVSGLESMPVFTPS 394

RESULT 6  
YJTB\_BACSU STANDARD; PRT; 396 AA.  
AC 034374;  
DT 15-FEB-2000 (Rel. 39, Created)  
DT 15-FEB-2000 (Rel. 39, Last sequence update)  
DT 15-FEB-2000 (Rel. 39, Last annotation update)  
DE PUTATIVE CYTOCHROME P450 YJTB (EC 1.14.-.-).  
GN YJTB  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.

RN [1]  
SEQUENCE FROM N.A.  
RC STRAIN=168;  
RA Rivoita C.; Soldo B.; Lazarevic V.; Joris B.; Manuel C.; Karamata D.;  
RT "A 35.7 kb DNA fragment from Bacillus subtilis chromosome containing a  
RT putative 12.3 kb operon involved in hexonulate catabolism and a  
RT perfect catabolite-responsive element."  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AF015825; AAC46317.1; -.  
CC DR EMBL: Z99110; CAB13078.1; -.  
CC DR SUBTILIST; BG13195; YJTB.  
CC PFAM: PF00067; P450; 1.  
CC PRINTS: PR00359; BP450.  
CC DR PROSITE: PS00086; CYTOCHROME\_P450; 1.  
CC DR Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.  
CC KW BINDING 349 349 HEME (BY SIMILARITY).  
CC FT BINDING 349 349 HEME (BY SIMILARITY).  
CC SQ SEQUENCE 396 AA; 44990 MW; 9A89CFL2613DBC6B CRC64;

Query Match 16.3%; Score 355; DB 1; Length 396;  
Best Local Similarity 30.1%; Pred. No. 1; 1e-19;  
Matches 99; Conservative 54; Mismatches 158; Indels 18; Gaps 7;

QY 71 IREAYEDYRHFSCEPTIPREAGENAYDPTSMPPROFALANQVGVVVDKLENR 130  
DB 56 VKKVVGDKELFSS-C--MPOQTSSIGNSI-INMDPPKHTKIRSVNKAFTPRVAKQMEPR 111  
QY 131 IOELACSLIESLRPOGOCNFTEDYAEPPIRIFMLAGLEEDIPHLKYLTDMFT----- 185  
DB 112 IOETIDELIQFGQSEFDLVHDSYPLVIVISELGLVPSAMHEQFRKMSDILVSTPKD 171  
QY 186 -RPDGSMTFAE---AKEALYDYLPIIEQRQKPGTDAISIVANGVNGRPTISDEAKR 240  
DB 172 KSEAEKAPFLIEBKDCBELLAAPFAGITIEKRNNPEQDIISILVEAETGKLSGEELIP 231  
QY 241 MFGLLVGLDVTNVPLSFMEFLAKSPENRQELIERPERIPACEELLR-RESLVADGR 299  
DB 232 FCTLLVAGNETTNNLISNAYSILETPGVYEELRSHPELMPQAVEBALRFRAPAVLRR 291  
QY 300 ILNSDYEFHVQQLKGDQIILPQMLSGIDERNKACPMHVDPSRKVSHTTGSHGLCLG 359  
DB 292 IAKRDTIEGHLIKEGMVLAFAVANRDEKFEEDRPHMFDIRRPHNPHIAGHIFCLG 351  
QY 360 QHLARREIYTLKEWLTTRIPDF--SIAP 385  
DB 352 APLARLEANIALISLSAFPMECVSTIP 380

RESULT 7  
CPXM\_BACSU STANDARD; PRT; 405 AA.  
AC P27632;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 15-FEB-2000 (Rel. 39, Last annotation update)  
DE CYTOCHROME P450 109 (EC 1.14.-.-) (ORF405).  
GN CYP109.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
RN [1]  
RC SEQUENCE FROM N.A.  
RC STRAIN=M23;





Db	403	V-VATEPBARLLSGFHGKTLPTVM	427
RESULT	9		
ID	CPXF-STRCO	STANDARD	PRT: 402 AA.
AC	P18327:		
DT	01-NOV-1990 (Rel. 16, Created)		
DT	01-NOV-1990 (Rel. 16, Last sequence update)		
DT	15-DEC-1998 (Rel. 37, Last annotation update)		
DE	CYTCHROME P450-SU2 (EC 1.14.-.-) (P450-CVB1) (CYP105B1).		
GN	CYP105B1 OR SUBC.		
OS	Streptomyces griseolus.		
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;		
OC	Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.		
RN	[1]		
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 1-32.		
RC	STRAIN-ATCC 11796;		
RX	MEDLINE: 90264332.		
RA	Omer C.A., Lenstra R., Little P.J., Dean C., Tepperman J.M.,		
RA	Leico K.J., Romesser J.A., O'Keefe D.P.;		
RT	"Genes for two herbicide-Inducible cytochromes P-450 from		
RT	Streptomyces griseolus.";		
RL	J. Bacteriol. 172:3335-3345(1990).		
CC	-1- FUNCTION: METABOLISM OF A NUMBER OF SULFONYUREA HERBICIDES.		
CC	-1- INDUCTION: BY HERBICIDE.		
CC	-1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.		
CC	-----		
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its use		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> ).		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
DR	EMBL: M32239; AAK26825.1; -		
DR	PIR: B35401; B35401.		
DR	HSSP: P23295; 2ROM.		
DR	PFAM: PF00067; P450.1.		
DR	PRINTS: PR00359; BP450.		
DR	PROSITE: PS00086; CYTOCHROME_P450.1.		
KW	Oxidoreductase; Monooxygenase; Electron transport; Heme.		
FT	INIT MET 0		
FT	BINDING 351 351 HEME (BY SIMILARITY).		
SO	SEQUENCE 402 AA; 44278 MW; E3B6F6F72C26D9D CRC64;		
Query Match	15.6%; Score 340; DB 1; Length 402;		
Best Local Similarity	28.9%; Pred. No. 1.6e-18;		
Matches 105; Conservative	63; Mismatches 165; Indels 30; Gaps 13.		
QY	63	WIATRGQIREAYEDYRHFSSSEC-----PFIPEGAEAYDIPT--SMDPEQROFRALA	115
DB	51	WLVTNRHODVRAVLGDPF-FSADNARHTGPFPLTAGGREITIGTPTFLRMDDPEHARLRML	109
QY	116	NOVGMPVVDKLENIQGLACSLIISLR-QGQCFTEDYAPFPPIRIFMLAGLPEED	174
DB	110	TADETVKKVEARPEVQRIADLVLRMTGTRISADLVTEFALPELSVLCILLGVYEDH	169
QY	175	PHL-----KYLTDQMRPQSGMTFAEAKELAYDLPIIEORQKQGTDAIS-IVANGQV	229
DB	170	AFQORSVLLTLRSTPE---EVRAQDELEYLARLARTKRRDDAIIISLVARGEID	226
QY	230	GRPITSDEAKRMFGILLVGGDTVNVFLSFSMEFLAKSPHROELIERERIPAAACELL	289
DB	227	DQOINT-----MGRLLVAGHETTNMTALSTVLVLRNDQARLRAPALYKGAVEELL	281
QY	290	RFFSIVADG--NILSDVEFHGVOLKKGQILLPQMLSGIDERNACP--MHVDSRKQV	345
DB	282	RLYLVHNGVPRPIATEDVLIGRRTIAGGQVLC--MISSANNDAEVPEGGDDLVAARDR	339

```

OY 346 SHTTGHSHTLGGHRLARREITVLKEMLTPEPFSA-PEAOIQHSGI-VSGVOLP 403
DB 340 RHVAFGEVHQGLPRLARVELQIAETLRLPDLRLAVPHEEITPFGDMATIGVHSLP 399
OY 404 LVW 406
DB 400 IAW 402

RESULT 10
CPXL_BACME STANDARD: PRT: 410 AA.
AC PI4762:
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CYTOCHROME P450(BM-1) (EC 1.14.14.1).
GN CYP106.
OS Bacillus megaterium.
OC Bacteria: Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A. AND SEQUENCE OF 1-25.
RC STRAIN-ATCC 14581;
RX MEDLINE: 90089408.
RA He J.S., Ruettinger R.T., Liu H.-M., Fulco A.J.;
RT "Molecular cloning, coding nucleotides and the deduced amino acid
RT sequence of P-450BM-1 from Bacillus megaterium.";
RL Biochim. Biophys. Acta 1009:301-303(1989).
RN [2]
RP SEQUENCE OF 1-25 FROM N.A.
RX MEDLINE: 95355495.
RA He J.S., Liang Q., Fulco A.J.;
RT "The molecular cloning and characterization of BM1P1 and BM1P2
RT proteins, putative positive transcription factors involved in
RT barbiturate-mediated induction of the genes encoding cytochrome
RT P450BM-1 of Bacillus megaterium.";
RL J. Biol. Chem. 270:18615-18625(1995).
CC -I- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
CC MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED
CC COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.
CC -I- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation --
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
-----
DB EMBL: X16610; CAA34612.1; -
DR EMBL: S79230; AAC60495.1; -
DR PIR: S07764; Q4BS6M.
DR HSSP: P33006; 1CPT.
DR PFAM: PF00067; p450. 1.
DR PRINTS: PR00359; BP450.
DR PROSITE: PS00086; CYTOCHROME_P450. 1.
DR Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.
KW BINDING 356 HEME.
FT SEQUENCE 410 AA; 47460 MW; C9AE293E76745387 CRC64;
SQ

Query Match 15.4%; Score 334.5; DB 1; Length 410;
Best Local Similarity 27.7%; Pred. No. 4.2e-18;
Matches 96; Conservative 57; Mismatches 170; Indels 23; Gaps 7;

OY 116 NOVGMPPVADKLNRNIQELACLIESLRPOGOCNFTEDYAEPPFIRIFMLAGLEEDIP 175
DB 45 MNVPDYENHKVQLSNYDPFSSDGOCTTTFVGDNSKKKSTPTNLNLTPDPHRRARSL 104

```

Db 105 AAFPRSLKNWEPRIKQIADLVEALQKNSTINIVDDLSPPPSVIADLFGVYKDRY 164  
 QY 176 HLKYLTDQMTRPDGSMTFAEAKE-----ALXYDLPIIEORROKPGTATISVANGOV 228  
 Db 165 QFKKWVDILFQRPDQERLEIEDEKORAGAEYQYLYPIYIEKRSNLSDDLISDLIAEV 224  
 QY 229 NGRPTISDEAKRMFGLLVGLDVTNVL-SFSMEFLAKSPREHROELIERPERIPAAEE 287  
 Db 225 DGEFTDEEIVHAATMLLGACVETTSIAINMEYSPFLYDOKSLYSLELRNRELAPKAVEE 284  
 QY 288 LLR-RPSLVADGRHLTSDYEFHGVOALKKGQDILLPQMLGLDERKNACPMHVDPSR-QKV 345  
 Db 285 MLKRFHISRRDTRVAKODNELLGVKLGKGDVVIAMMSACNMDETFMENPSPVDIHRPTNK 344  
 QY 346 SHTFGHSGLCGOHLARREITVTLKEMLTR-----IPDSIAP 385  
 Db 345 KHLTFGGRPHFCIGAPLARLEMKLLEAFLEAFSHTEPFEDFELEP 390

RESULT 11  
 CPXY\_BACSU STANDARD: PRT: 410 AA.  
 AC 008469:  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-FEB-2000 (Rel. 39, Last annotation update)  
 DE CYTOCHROME P450 (EC 1.14.-.-).  
 GN CYP4 OR CYP107J1.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168 / BGSC1A1;  
 RX MEDLINE: 97431495.  
 RA Belitsky B.R., Gustafsson M.C.U., Sonenshein A.L., von Wachenfeldt C.;  
 RT "An lrp-like gene of Bacillus subtilis involved in branched-chain  
 RT amino acid transport.";  
 RL J. Bacteriol. 179:5448-5457(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE: 97453479.  
 RA Sorokin A., Bolotin A., Purnelle B., Hilbert H., Lauber J.,  
 RA Duesterhoeft A., Ehrlich S.D.;  
 RT "Sequence of the Bacillus subtilis genome region in the vicinity of  
 RT the lrp operon reveals two new extracytoplasmic function RNA  
 RT polymerase sigma factors Sigv and Sigz.";  
 RL Microbiology 143:2939-2943(1997).  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Y11043; CAAT1937.1; -  
 DR EMBL: U93876; AAB80898.1; -  
 DR EMBL: Z59117; CAB14615.1; -  
 DR HSSP: Q00441; IOXA.  
 DR SUBTLIST: BG11929; CYP4.  
 DR PIR: P00067; P450.1.  
 DR PRINTS: PR00359; BP450.  
 DR PROSITE: PS00086; CYTOCHROME\_P450; 1.  
 DR OXIDOREDUCTASE: Monooxygenase; Heme.  
 FT BINDING 359 359 HEME (BY SIMILARITY).  
 FT SEQUENCE 410 AA; 47384 MW; 035D98B58A4761AB CRC64;

Query Match 15.3%; Score 334; DB 1; Length 410;

Best Local Similarity 27.2%; Pred. No. 4, 6e-18;  
 Matches 86; Conservative 69; Mismatches 139; Indels 22; Gaps 5;  
 QY 104 DPEQOFALANOVGMVYDLENRIQELACSLIESLRPOQOCFTDYAPRIRLF 163  
 Db 100 DPEQHNRKRLTVOKAFTHRTTHRIQLDLEKTIQIHLADSKVQPKMKMLVDDYAPRIRIY 159  
 QY 164 MLTAGPREEDIPHLKYLTDQMTRPDGSMTFAEAKEL-----YDLPIIEQRQ 213  
 Db 160 SEMGLPLEDRQKFRVWSQAI-----IDFSAPRELGQENDHLGFEVLESLVKKRR 213  
 QY 214 KPGTDAISVANGVNGRPRTISDEAKRMFGLLVGLDVTNVLFSFSMEFLAKSPREHROE 273  
 Db 214 EPAGDILSALIQASEGTOISTEELYSMILILVAGHETVTNLTMTFYALMCHNDQLEK 273  
 QY 274 LIERPERIPACEELLRRSLV--ADGRILTSDYEFHGVOALKKGQDILLPQMLSGDERK 331  
 Db 274 LRQOPDLMSALIEALREHSPVLETTIRMTAEPRFLLHGOEIKRKDYIIISLASANDERK 333  
 QY 332 NACPMHVEDSRQKVSHTTFGHSGLCGOHLARREITVTLKEMLTRIPPSIAPGAQIOH 391  
 Db 334 FRPADIFDIERKKNRHLIACGHGHPCLGAGLAPLEKIALISTLLRCRPNQL-KGEKKOM 392  
 QY 392 K---SGIVSGVALPL 404  
 Db 393 KMGKFMRLALELPL 408

RESULT 12  
 CPXY\_SACER STANDARD: PRT: 405 AA.  
 AC P33271;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE CYTOCHROME P450 107B1 (EC 1.14.-.-) (P450C11B1).  
 GN CYP107B1.  
 OS Saccharopolyspora erythraea (Streptomyces erythraeus).  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Pseudonocardineae; Pseudonocardaceae;  
 OC Saccharopolyspora.  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 172-203 AND 302-321.  
 RC STRAIN=NRRL 2338;  
 RX MEDLINE: 92121109.  
 RA Andersen J.F., Hutchinson C.R.;  
 RT "Characterization of Saccharopolyspora erythraea cytochrome P-450  
 RT genes and enzymes, including 6-deoxyerythronolide B hydroxylase.";  
 RL J. Bacteriol. 174:725-735(1992).  
 CC -1- FUNCTION: NOT KNOWN, PROBABLY INVOLVED IN THE CATABOLISM OF  
 CC OCTANE AND GUAIACOL. IT DISPLAYS A WEAK ACTIVITY IN THE O-  
 CC DEALKYLATION OF 7-ETHOXYCOUMARIN.  
 CC -----  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M83110; AAA26483.1; -  
 DR PIR: BA2606; B42606.  
 DR HSSP: Q00441; IOXA.  
 DR PIR: P00067; P450.1.  
 DR PRINTS: PR00359; BP450.  
 DR PROSITE: PS00086; CYTOCHROME\_P450; 1.  
 DR OXIDOREDUCTASE: Monooxygenase; Electon transport; Membrane; Heme.  
 FT BINDING 352 352 HEME (BY SIMILARITY).  
 FT SEQUENCE 405 AA; 45238 MW; 71C93CEC1DC53FD CRC64;





SQ SEQUENCE 405 AA; 44580 MM; 72DEAEC6B688FA48 CRC64;

Query Match 14.6%; Score 318.5; DB 1; Length 405;

Best Local Similarity 29.5%; Pred. No. 6,7e-17;  
Matches 102; Conservative 51; Mismatches 160; Indels 33; Gaps 9;

```
QY 62 HMIATRGQLIREAYEDYRHFSECPPIPREAGE----AYDFIPTSN-DPEQORFRALA 115
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 46 YYLSRSHADYWSAARDHQFESS-AQGLTVNYGLEMIGLHDTPPMWQDPPVHTFERKLY 104
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 116 NQYVGMPEVYDKLENRIQELACSLIESLRPOGOCNFTEDYNAEPPPIRIEMLLAGLPEEDIP 175
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 105 SRGFTPROVETVEPTVRKFVEVERLEKLRANGGSDIVTELFKPLPSMVVAHYLGVPPEEDMT 164
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 176 HLKYLTDQMRP---DGSMTFA-EAKEALDYLLPIIEORROKPGTDATIS--IVANGQVN 229
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 165 QPDGWTQAIYANANAVDGAATTGALDAVGSMMAYFTGLTERRRTEPADDAISHLVNAGVGAD 224
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 230 GRPITSDEAKRMFGLLVGGLDVTVNFLSFSMEFLAKSPEROELTERPERIPACEELL 289
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 225 GDTAGTLSILAFTFTWVTGNDVTGMLGSMFLHRRPDQRLLDDPREGIPDAVEELL 284
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 290 RRFSLYAD-GRILTSYEFHGVQKKGDQIILLPQMLSGLDERK-----NACPMHV 338
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 285 RLTSPTYQGLARTTRDTYITIGDTTIPAGRRLVLLYGSANRDERQYGPDAELDVTRCPRNI 344
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 339 DFSROKVSHTTFGHSHLCLGQHLARREIIVTLKEWLTIRIPDFSIA 384
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 345 -----LTFSHGAHHCIGAAAAARMCRAVLTLLARCPDFEVA 381
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

Search completed: October 4, 2000, 13:04:55  
Job time: 1686 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 12:39:37 ; Search time 79.83 Seconds

(without alignments)  
359,568 Million cell updates/sec

Title: US-09-246-451-13

Perfect score: 2177  
Sequence: 1 TTERIQSNANLAPLPPIHVE.....IVSGVQALPLVWDPATTKAV 414

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SPREMBL\_12:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mammal:\*  
8: sp\_mhc:\*  
9: sp\_organeller:\*  
10: sp\_phage:\*  
11: sp\_plant:\*  
12: sp\_rodent:\*  
13: sp\_virus:\*  
14: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	364	16.7	396	2	052544 amycolatops
2	363.5	16.7	416	2	087605 streptomyc
3	362	16.6	397	2	059523 micromonosp
4	345	15.8	410	2	09XSP9 streptomyc
5	344.5	15.8	404	2	054302 streptomyc
6	344.5	15.8	407	2	059819 streptomyc
7	337	15.5	411	2	032460 actinomadur
8	327	15.0	406	2	059723 pseudomonas
9	325.5	15.0	376	2	031785 bacillus su
10	323.5	14.9	420	2	09ZHQ1 streptomyc
11	321	14.7	407	2	09XSP8 streptomyc
12	315	14.5	406	2	087192 streptomyc
13	314	14.4	410	2	09XSP7 streptomyc
14	313	14.4	400	2	09XDB0 mycobacteri
15	311.5	14.3	410	2	059831 streptomyc
16	307	14.1	310	2	032927 mycobacteri
17	304.5	14.0	397	2	024727 nocardioid
18	303	13.9	417	2	059910 streptomyc
19	299	13.7	388	2	P77977 streptomyc

20	299	13.7	406	2	085697 streptomyc
21	291	13.4	351	2	052572 amycolatops
22	283.5	13.0	337	2	052816 amycolatops
23	280.5	12.9	406	2	087675 amycolatops
24	280	12.9	421	2	052561 amycolatops
25	272	12.5	403	2	052560 amycolatops
26	269.5	12.4	386	2	059921 streptomyc
27	266.5	12.2	310	2	P96562 amycolatops
28	263.5	12.1	437	2	09ZPC0 mycobacteri
29	262	12.0	395	2	09XST2 streptomyc
30	256	11.8	398	2	087674 amycolatops
31	254	11.7	426	2	09ZFC3 mycobacteri
32	251.5	11.6	411	2	09X8Q3 streptomyc
33	250	11.5	411	2	060005 putative
34	249.5	11.5	511	2	052569 amycolatops
35	244	11.2	406	2	052823 amycolatops
36	240	11.0	433	2	09XC6 streptomyc
37	225	10.3	398	2	059079 amycolata a
38	224	10.3	391	2	052822 amycolatops
39	222.5	10.2	391	2	087673 amycolatops
40	204.5	9.4	419	2	050242 agrobacteri
41	204	9.4	397	2	052802 amycolatops
42	182.5	8.4	119	2	085655 streptomyc
43	179	8.2	120	2	085653 streptallo
44	172.5	7.9	313	2	09X4I8 myxococcus
45	164	7.5	120	2	085650 streptomyc

## ALIGNMENTS

RESULT 1  
ID 052544 PRELIMINARY; PRT; 396 AA.  
AC 052544;  
DT 01-JUN-1998 (TREMblrel. 06, Created)  
DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)  
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)  
DE CYTOCHROME P450 MONOOXYGENASE.  
OS Amycolatopsis mediterranei.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Amycolatopsis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SRRAIN-S699;  
RA KIM C.G., YU T.W., FRYHLE C., HANDA S., FLOSS H.G.;  
RL J. Biol. Chem. 270:1998.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SRRAIN-S699;  
RA AUGUST P.R., TANG L., YOON Y.J., NING S., MUELLER R., YU T.W.,  
RA TAYLOR M., HOFFMANN D., KIM C.G., ZHANG X., HUTCHINSON C.R.,  
RA FLOSS H.G.;  
RL Chem. Biol. 5:0-0(0002).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SRRAIN-S699;  
RA AUGUST P.R., TANG L., YOON Y.J., NING S., MUELLER R., HUTCHINSON C.R.,  
RA FLOSS H.G.;  
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF040570; AAC01709.1; -;  
DR HSSP: 000441; 10XA.  
DR PRAM: PR00067; P450.1.  
DR PRINTS: PR00359; BP450.  
KW Monooxygenase.  
SQ SEQUENCE 396 AA: 44293 MW: BA875129 CRC32:

Query Match 16.7%; Score 364; DB 2; Length 396;  
Best Local Similarity 29.08; Pred. No. 1.9e-21;  
Matches 117; Conservative 59; Mismatches 183; Indels 44; Gaps 14;  
QY 27 DWYPSNLASGVQENAVLVQESNVPLVWTRCNGH--WIATRCQLITREAYEDYRHSSE 84

```

Db 15 DKEDPAVEDSLKEE-RPLAKMVPD-----GHGWMVSSYSELVREVLSDLR-FSHS 64
QY 85 CP---FIPREAGVADPIPT-----SMDPEQROFALANOVYGMVVDKLENRIQ 132
Db 65 CEVGHFPPVHQGV---IPRHLIPGMFTIMDPPEHTRRYKLLTGERTVVARASRLIPRAE 121
QY 133 ELACSLIESLRPG-OCNFTEDYAEPPRIIFMLAGLPEED---IPHLKYLTDOMTRP 187
Db 122 AVAAEEIEVVARAKADADVMDFAKPLVLEMLGELVCLPYEERDRVPAVATLLHDAEADP 181
QY 188 DSGMTFAEAKALYDLIPRIEORROKPGTDAISYANGVNRPTSDAKMFGLLY 247
Db 182 AEA---AAAEVAGKPFDEYIERRORPDDLISSLVT-----EDLTQELRNITVLLF 233
QY 248 GGLDTVNFISFSEMEFLAKSPENRQELIERPERIPACEELLRRFSL--VADGRILTSYD 305
Db 234 AGETTEGALATGVFALLHHTDQALAKPEPKLDAAIEELLKLYVNOYHTTRTALEDV 293
QY 306 EPHVOLKKGDOILLPOMLSGLDERKNACPMHYDFSROKVSHTTFGSHLGLGHLARR 365
Db 294 KLGELIKKGDVYVSLPAANRDPAKFCPAFLDIEDRTSGHVAFGHIGCLGONLARI 353
QY 366 EIIVTLKEMWLTIRPDPSIA-PGAQIOHK-SGLYSGVALPLW 406
Db 354 ETRAGFTALLRAPELRLAVPADEVPRLKGSVSVKKLPLVSM 396

RESULT 2
087605 PRELIMINARY: PRT: 416 AA.
AC 087605;
DT 01-NOV-1998 (TREMblrel, 08, Created)
DT 01-NOV-1998 (TREMblrel, 08, Last sequence update)
DT 01-NOV-1999 (TREMblrel, 12, Last annotation update)
DE CYTOCHROME P450 MONOOXYGENASE.
GN PICK OR PIKC.
OS Streptomyces violaceus (Streptomyces venezuelae), and
OC Streptomyces venezuelae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC15439;
RA BETLACH M.C., KEALEY J.T., BETLACH M.C., ASHLEY G.W., MCDANIEL R.;
RT "Characterization of the macroliide P450 hydroxylase from Streptomyces
RT venezuelae which converts narbomycin to picromycin.";
RL Biochemistry 0:0-0(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES-S.VENEZUELAE; STRAIN=ATCC15439;
RX MEDLINE: 99051447.
RA XUE Y., WILSON D., ZHAO L., LIU H.-W., SHERMAN D.H.;
RT "Hydroxylation of macroliactones YC-17 and narbomycin is mediated by
RT the PIK-encoded cytochrome P450 in Streptomyces venezuelae.";
RL Chem. Biol. 5:661-669(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES-S.VENEZUELAE; STRAIN=ATCC15439;
RX MEDLINE: 98445333.
RA XUE Y., ZHAO L., LIU H.-W., SHERMAN D.H.;
RT "A gene cluster for macroliide antibiotic biosynthesis in Streptomyces
RT venezuelae: architecture of metabolic diversity.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:12111-12116(1998).
DR EMBL: AF087022; AAC64105.1; -.
DR EMBL: AF079139; AAC68886.1; -.
DR HSSP: Q00441; IOXA.
DR PFAM: PF00067; P450.1.
DR PRINTS: PR00359; BP450.
KW Monooxygenase.
SQ SEQUENCE 416 AA; 46038 MW; CAD6EBA0 CRC32;

```

```

Query Match 16.7%; Score 363.5; DB 2; Length 416;
Best Local Similarity 28.3%; Pred. No. 2.3e-21;
Matches 97; Conservative 62; Mismatches 163; Indels 21; Gaps 7;

QY 76 EDYRHSSECPPIPRAGEYDIFPTSMDDPEORORALANOVYGMVVDKLENRIQELA 135
Db 72 KDMR--NSTPTLEAENALHNHNL--DEPRHTRLRKLVAREFTRRRELLRPROVEIV 127
QY 136 CSLIESL--RPGOCNFTEDYAEPPRIIFMLAGLPEEDIPHLKYLTDOMTRPGSMF 193
Db 128 DGLVDMALAAPDGRADLMESLAMPRLITVISELGYPEPDRAPFRWTDFAFVPPDPAQ 187
QY 194 AEAKALYDLIPRIEORROKPGTDAIS-IVANGVNRPTSDAKMFGLLVGLDPT 252
Db 188 QTAAEMSGYLSRLIDSKRQDEDELLSALVRTSDBDGSRLTSEELGMHILLVAGHET 247
QY 253 VVNFISFSEMEFLAKSPENRQELIERPERIPACEELLRRFSLVADRILTSYEF----- 307
Db 248 TVNLIANGVALLSHPDQALALRADMTLLDGAVEMLR---YEBPVRSATYRFVEEYV 302
QY 308 --HGVOLKKGDOILLPOMLSGLDERKNACPMHYDFSROKVSHTTFGSHLGLGHLARR 365
Db 303 DLDGYIIPADYLVVYLAADAHRTPERPPRPHRDIRROTAGHLAFGHGHCIGAPLARL 362
QY 366 EIIVTLKEMWLTIRPDPS--IAPGAQIOHKSGLYSGVALPLW 406
Db 363 EARIARALLERCPLDALDVSPGELVWYPMPIRGKALPIRW 405

RESULT 3
059523 PRELIMINARY: PRT: 397 AA.
AC 059523;
DT 01-NOV-1996 (TREMblrel, 01, Created)
DT 01-NOV-1996 (TREMblrel, 01, Last sequence update)
DT 01-NOV-1999 (TREMblrel, 12, Last annotation update)
DE CYTOCHROME P450-LIKE PROTEIN (EC 1.14.14.1).
GN MYC.
OS Micromonospora griseorubida.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micromonosporinae; Micromonosporaceae;
OC Micromonospora.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A11725;
RX MEDLINE: 95107242.
RA INOYE M., TAKADA Y., MUTO N., HORINOCHI S., BEPPU T.;
RT "Characterization and expression of a P-450-like mycinamicin
RT biosynthesis gene using a novel Micromonospora-Escherichia coli
RT shuttle cosmid vector.";
RL Mol. Gen. Genet. 245:456-464(1994).
DR EMBL: D16098; BAA03672.1; -.
DR HSSP: Q00441; IOXA.
DR PFAM: PF00067; P450.1.
DR PRINTS: PR00359; BP450.
SQ SEQUENCE 397 AA; 44331 MW; DID592A6 CRC32;

```

```

Query Match 16.6%; Score 362; DB 2; Length 397;
Best Local Similarity 28.5%; Pred. No. 2.8e-21;
Matches 105; Conservative 57; Mismatches 168; Indels 38; Gaps 8;

QY 63 WATRQQLRBEAYEDYRHHSSSECPFI-----PRAGEAVDFIPTSMDDPEORQFRA 113
Db 44 WLAVTR-----YEDVRAVLGGRFVRGSPSMTRDEPRTREPENKGGLSMDPESHRLRR 96
QY 114 LANOVYGMVVDKLENRIQELACSLIESLRPGQ-CNFTEDYAEPPRIIFMLAGLPEE 172
Db 97 LVYKATFARASLSIRRAEINHELVDQMAATGQPADLVAMARQOLPVAIVICELGVPSA 156
QY 173 DIPHLKYLTDOMTRPDGSM-----TPAEAKALYDLIPRIEORROKPGTDAISIV 223
Db 157 D-----HDFRFRMGAGPLSTAENVTAEMQEAAGVAYVMDLLIDRRRKEPTDVLVAL 209

```



```
OY 224 ANGVNGRPTSDAKRMGLLVGGLTVVNF.SFSMEFLAKSPHRELLERPERIPA 283
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 210 VQARDODDSISEGELLDLAIGLLVAGYESTPTQIADFYVLLMTRELRQLDRPELIPS 269
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 284 ACEELLRRPSL--VADGRLLTSDYEFHGYOLKKQDQILLPOMLSGLDERKNACPMHVD 340
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 270 AVEELTRWVPLGVGTAFPPRAVEDVTLRGVTTIRAGSEVPLASTGANNDQAFDADRIDV 329
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 341 SRQVSHHTFGHSHCLGSHLARRELLITLTKEMLRIPDPSIA--PGAQIOHKS--IYSG 398
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 330 DRTPNHGLGSHGHCLGAPLARVELQALVLELLQRLPGLRIGIPETQURWSEGMILRG 389
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 399 VQALPLVW 406
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 390 PLELPVW 397

RESULT 4
OY5P9 PRELIMINARY: PRT: 410 AA.
AC OY5P9:
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE CYTOCHROME P450 HYDROXYLASE ORF3.
OS Streptomyces lavendulae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 2564;
RX MEDLINE: 99201491.
RA MAO Y.Q., VAROGLU M., SHERMAN D.H.;
RT "Molecular characterization and analysis of the biosynthetic gene
RT cluster for the antitumor antibiotic mitomycin C from Streptomyces
RT lavendulae NRRL 2564."
RT Chem. Biol. 6:251-263(1999).
DR EMBL: AF127374; AAD28449.1;
SQ SEQUENCE 410 AA; 45887 MW; 507C7F38 CRC32;
```

Query Match 15.8%; Score 345; DB 2; Length 410;  
Best Local Similarity 27.0%; Pred. No. 6.8e-20;  
Matches 113; Conservative 69; Mismatches 193; Indels 44; Gaps 14;

```
OY 9 ANIAPLRP-HVPHLVDEDMWPSNLSAGVQDAMAVLOESNVPLVTRCNGCHWIATR 67
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 15 AGEPAVYPPFHAPDRL--EPDPY-----WEPLRRERPLQRTVTLTPYGGSEWLATR 60
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 68 GQIIRAEYEDYRHFSSE---CFPIREAGRAVFIPI-----TSMDDPEQOQFALAN 116
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 61 YQDVRAVFAAD-RFESQQLAVAPGAR-----FLPHQPPDVALVEGDHARLRRLVG 112
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 117 QVQVMFVNDKLEARIQELACSLIESLRPOG-OCNFTEDYAEPPIRIPLMLAGLPEEDIP 175
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 113 KVTPTRRVEDMERLLQRTADGLDAMEKMPRADLYEDPSLPRFAVSMICGLLVPEPDRK 172
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 176 HUKYLTDMQTRPDGSMTPFAAKE--ALVDYLPIITIQRRQKPGTDAISIVANGVNGRP 232
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 173 RFCVWDALTLTTT-AHTPAQVRDYMQMMDYLGGLVAQRRVPTADLIGSLVRADEEDK 231
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 233 ITSDAKRMFGLLVGGGLDVVNF.SFSMEFLAKSPHRELLERPERIPACCELLRRF 292
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 232 LVEGELVRLAEALLIAGYETSAQIPIFVLFRHPOLLERIRINDHDLPAVAEELLRFV 291
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 293 SL-VADG--RIILTSDEYFNGVOLKKQDQILLPOMLSGLDERKNACPMHVDPSROKVSHTT 349
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 292 PICTVGGFPRTATEDVELGVLVRAGETVVPVSGAANRDELLTDPDELDLARRPNHLG 351
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 350 FGGSHCLGQHLARREITVTLKEMLRIPDPSIA--PGAQIOHKS--IYSGVQALPLVW 406
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 352 FGAGRHHCIGAOIARVELQITLTTLFRRYPRRLAVPEESLSWKEGILMVGMMHTMPTVW 410
```

```
RESULT 5
OY54302 PRELIMINARY: PRT: 404 AA.
AC OY54302:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE CYTOCHROME P450.
GN RAPN.
OS Streptomyces hygroscopicus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 5491;
RX MEDLINE: 95372374.
RA SCHWECKE T., APARICIO J.F., MOLNAR I., KOENIG A., KHAW L.E.,
RA HAYDOCK S.F., OLIVNYK M., CAFFEY P., CORTES J., LESTER J.B.,
RA BOEHM G.A., STAUNTON J., LEADLAY P.F.;
RT "The biosynthetic gene cluster for the polyketide immunosuppressant
RT rapamycin."
RT Proc. Natl. Acad. Sci. U.S.A. 92:7839-7843(1995).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 5491;
RA MOLNAR I., APARICIO J.F., HAYDOCK S.F., BE KHAW L., SCHWECKE T.,
RA KOENIG A., STAUNTON J., LEADLAY P.F., STAUNTON J., LEADLAY P.F.;
RL Gene 0:0-0(0).
RN (3)
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 5491;
RA APARICIO J.F., MOLNAR I., SCHWECKE T., KOENIG A., HAYDOCK S.F.,
RA BE KHAW L., STAUNTON J., LEADLAY P.F., LESTER J.B., BOEHM G.A.,
RA STAUNTON J., LEADLAY P.F.;
RL Gene 0:0-0(0).
DR EMBL: X86780; CAA60465.1;
DR HSSP: 000441; IOXA.
DR PFAM: PF00067; P450; 1.
DR PRINTS: PR00359; BP450.
SQ SEQUENCE 404 AA; 45071 MW; 05AB94DF CRC32;
```

Query Match 15.8%; Score 344.5; DB 2; Length 404;  
Best Local Similarity 28.6%; Pred. No. 7.3e-20;  
Matches 102; Conservative 63; Mismatches 177; Indels 15; Gaps 9;

```
OY 63 WINTRGQIAREAYEDYRHFSSECF--IPRAGEAYDF-IPTSMDPPEOROFALANOV 118
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 50 WLAASMEDVAKVVDPR-FERSATLKGQDVPRVLPAIDQOVYIMLDPPELTRLRVATKA 108
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 119 VGMFVNDKLEARIQELACSLIESLRPOG-OCNFTEDYAEPPIRIPLMLAGLPEEDIPHL 177
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 109 LTSRRMEALRPFOEVADDLIDMKLAGAPADLMEDPALPLPIMICELGVPLEDQTKF 168
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 178 KYLTDMQTRPDGSMTP--PAEAKALDYLIPIITIQRRQKPGTDAISIVANGVNGRPIT 234
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 169 RTWSDQML-SNGAYSQEVNAAQOSLYLISELIAERKKQDITDLGSLVRAHDKDRLS 227
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 235 SDAKRMFGLLVGGGLDVVNF.SFSMEFLAKSPHRELLERPERIPACCELLRRFSL 294
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 228 ETELVGFAVTLIAGETTNAINGSYTLTLTPREKIAELRKDLSLIPKAVDELRLTPIPI 287
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 295 VADG--RIILTSDEYFNGVOLKKQDQILLPOMLSGLDERKNACPMHVDPSROKVSHTT 351
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 288 AKQASWVMAVEDVELSGTIVKAGEAVALIOTHSANTDPKYYDHPREDDFRTSNPHMSLG 347
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 352 HGGSHCLGQHLARREITVTLKEMLRIPDPSIA--PGAQIOHKS--IYSGVQALPLVW 406
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 348 HGAHHCQADLVVENQOTALGSLISRIPLARFAVPEPRIRKIFRGRLVPSLEALPLW 404
```

RESULT	6			
059819	ID	Q59819	PRELIMINARY:	PRT: 407 AA.
AC		Q59819:		
DT		01-NOV-1996 (TREMBLrel. 01, Created)		
DT		01-NOV-1996 (TREMBLrel. 01, last sequence update)		
DT		01-NOV-1999 (TREMBLrel. 12, last annotation update)		
DE		CYTCHROME P450 (EC 1.14.14.1).		
GN		OLEP.		
OS		Streptomyces antibioticus.		
OC		Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;		
OC		Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.		
RN		[1]		
RP		SEQUENCE FROM N.A.		
RX		MEDLINE: 95255619.		
RA		RODRIGUEZ A.M., OLANO C., MENDEZ C., HUTCHINSON C.R., SALAS J.A.;		
RT		"A cytochrome P450-like gene possibly involved in oleandomycin		
RT		biosynthesis by Streptomyces antibioticus.";		
RL		FEMS Microbiol. Lett. 127:117-120(1995).		
CC		-I- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.		
DR		EMBL: L37200; AA92553.1; ..		
DR		HSSB: Q0041; I0XA.		
DR		PROSITE: PS00086; CYTOCHROME_P450; 1.		
DR		PFAM: PF00067; P450; 1.		
DR		PRINTS: PR00359; BP450.		
KW		Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.		
FT		BINDING 356 HEME (BY SIMILARITY).		
SO		SEQUENCE 407 AA; 44957 MW; 04411C60 CRC32;		

[illegible]

OC Actinomycetales; Streptosporangineae; Thermomonosporaceae;  
OC Actinomadura.  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=PI57-2;  
RX MEDLINE; 97480928.  
RA DAIRI T., HAMANO Y., IGARASHI Y., FURUMAI T., OKI T.;  
RT "Cloning and nucleotide sequence of the putative polyketide synthase  
RT genes for piridinichin biosynthesis from Actinomadura hibiscus.";  
RL Biosci. Biotechnol. Biochem. 61:1445-1453(1997).  
DR HSSL; D87924; BAA23153.1; -.  
DR HSSP; O00441; 10XA.  
PR PPM; PF00067; P450.1.  
DR PRINTS; PR00359; BP450.  
SQ SEQUENCE 411 AA; 44860 MW; 98A628B6 CRC32;

Query Match	Score	DB	Length
Match 117: Conservative	27.08; Pred. No. 3e-19; Mismatches 196; Indels 60; Gaps 17;		
QY 4	TIOSNANLAPRPVPHPEHLVEFDYKYP-----SNLSAGVOEAMAVLQESNVPDLWTRCN 59		
DB 9	TYDPRDPTPAFPFRPD-----DPFGPPCEHARLRASDPVAKVLP-----T 50		
QY 60	GGH-WIATRGQILIRATYIDYRHFSEC---PIPRAGEAATDPIPT-----MDPE 107		
DB 51	GDHAWVATRYADVRFVTS-D-RFSEAVTRPGAP-----LIPMGRSKSLYIMDPE 102		
QY 108	QROFALANOVYGMGVNDYKLENRIQELCSLIESLRPOG-OCNFEDYAEPPPIRIFML 166		
DB 103	HTBRMKIVSRAPFTARRVCGMARHARDLTSGVDEVDEVEHGPADLIALHLPLPTVICM 162		
QY 167	AGLPEEDIPHLKYLTDQMPRDSKTEFAEKE-----ALYDYLIPIEORQRPGTDAI 220		
DB 163	LGVPEDPRPDQMDTRML-TIGAPALAOADEIKAAVGRGLYLAELDARTAPADLL 221		
QY 221	SIVANQVGNRPRTSDEKRMFGL-LVGGGLDTVYNFLSPSMEFLASPEHROELIERPE 279		
DB 222	SLSRAHADDG--LSEELLTFGMLTLAAGCHHTTAITHSYHLLRPSRYARLRDPS 279		
QY 280	RIPACEELLRRFSLVADG---RLTSDYDFHGVOLKKGDOILLPOMLSGLDERKNACP 335		
DB 280	GIPAAVEELL-RYGOIGSGAGAIRAVDEVGVGLVAGAVIPLFNAARDEVFADP 338		
QY 336	MHVESRQKSVHTTFGHGSHLCLGCHLARRELIYTKMKMLRIRDFSTA-PCAQIQKSG 394		
DB 339	EELDGRTDNPHIALGHGHCYGLAPLARLEQVLLVETLVERTPALRLAIDADITWRPG 398		
QY 395	IV-SGVQALPLW 406		
DB 399	LAFARPDALPIAW 411		
RESULT 8			
ID 059723	PRELIMINARY;	PRT;	406 AA.
AC 059723;			
DT 01-NOV-1996 (TREMblrel. 01, Created)			
DT 01-NOV-1996 (TREMblrel. 01, last sequence update)			
DT 01-NOV-1999 (TREMblrel. 12, last annotation update)			
DE CYTOCHROME P450 LIN (EC 1.14.14.1).			
GN LINC.			
OS Pseudomonas incognita.			
OC Bacteria; Proteobacteria.			
RN 11			
RA ROPE J.D., GUNSAUS I.C., SLIGAR S.G.;			
RL Submitted (Jan-1994) to the EMBL/GenBank/DBJ databases.			
-1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.			
DR EMBL; L23310; AAA25810.1; -.			
DR HSSP; P33006; ICPT.			
DR PROSITE; PS00086; CYTOCHROME P450; 1.			





Db 324 NRDSAKFDRDELDPRPTLRPHMTFGWGAHHCGLAPLATMELEVASSTLLTTRPALRLDV 383  
QY 384 APGAQIQHKSGIVGVALPLVW 406  
Db 384 PREDVSNMTSITMRYPLALPVTM 406

RESULT 13  
Q9X9P7 PRELIMINARY; PRT; 410 AA.  
AC Q9X9P7;  
DT 01-NOV-1999 (TReMBLrel. 12, Created)  
DT 01-NOV-1999 (TReMBLrel. 12, last sequence update)  
DE 01-NOV-1999 (TReMBLrel. 12, last annotation update)  
GN NIKE PROTEIN.  
OS Streptomyces tendae.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TRUE901;  
RA BRUNTNER C., LAUER B., SCHWARZ W., MOEHRL E., BORMANN C.;  
RT "Molecular characterization of co-transcribed genes from Streptomyces  
RT tendae Tu901 involved in the biosynthesis of the peptidyl moiety of  
RT the peptidyl nucleoside antibiotic nikkomycin.";  
RL Mol. Gen. Genet. 0:0-0(0).  
DR EMBL: Y18574; CAB46536.1; -  
SQ SEQUENCE 410 AA; 45884 MW; 9B52CB74 CRC32;

Query Match 14.4%; Score 314; DB 2; Length 410;  
Best Local Similarity 27.1%; Pred. No. 2.1e-17;  
Matches 98; Conservative 64; Mismatches 180; Indels 20; Gaps 9;

QY 63 WIATRGQLIREAYEDYRHFSSSEC-PRIP-----REAGEAYDFIP-----SMDPEQR 109  
Db 47 WLVLKHLARKLLADPRVSADRLHAPRGRULAEQPRATERVRRLSTRSMIHLDGDEHG 106  
QY 110 QFRALANOVVGMVVDKLENRIOELA-CSLIESLRPOGOCNFTEDYAEPPRIEMLLAG 168  
Db 107 AHKKILTGFESLRIRALRRVQELIDYRSIDEMLAPOPADVEHVSQAAPSLVICELG 166  
QY 169 LPREDIPHL-KYLLDQMTPRDGSMTFAEAKELDYDLPIL-BOOROKPGDAISYANG 226  
Db 167 VPHEDQRDFEMAGMLYSKSVSTREAAASDALNDFLEDLYTEKEKEPTDDLIGRLIAR 226  
QY 227 QVNGRPITSDAKRMFGLLVGGIDTVVNFLSFSMEFLAKSPENHOELIERPERIPACE 286  
Db 227 NRRTPVMTHEIYGTAVMLDIAGHQTANNISLGVALLENPEKKARIADPSLLPAIE 286  
QY 287 ELLRRSLV--ADGRILTSYEFHGVOLKKGDQILLPOMLSGLDERKNACPMHVESRQK 344  
Db 287 EMLRYSVVENAPARVATEDIEIGVATIRKDEGIIVSGLAADMDEVEFHRDLDFERGA 346  
QY 345 VSHTTGHSGLLGLGHLARREIIVTLKEMLTRIPDSIA-PCAQIQHKSGI-VSGVAL 402  
Db 347 RHNVAGSYGVHQCGLGMALHARLEIYEFETLLRRVPGSLAVPAEELRYKDDAGIYGIVY 406  
QY 403 PL 404  
Db 407 PV 408

RESULT 14  
Q9XDBO PRELIMINARY; PRT; 400 AA.  
AC Q9XDBO;  
DT 01-NOV-1999 (TReMBLrel. 12, Created)  
DT 01-NOV-1999 (TReMBLrel. 12, last sequence update)  
DE 01-NOV-1999 (TReMBLrel. 12, last annotation update)  
GN CYTOCHROME P450.  
PIPA.

OS Mycobacterium smegmatis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MC2155;  
RX MEDLINE; 99287823.  
RA POUJIN P., DUCROCQ V., HALTIER-SOULIER S., TRUFFAUT N.;  
RT "Cloning and characterization of the genes encoding a cytochrome P450  
RT regulatory protein (PipR) in Mycobacterium smegmatis mc2155.";  
RL J. Bacteriol. 181:3419-3426(1999).  
DR EMBL: AF102510; AAD28344.1; -  
SQ SEQUENCE 400 AA; 44747 MW; C10DD01A CRC32;

Query Match 14.4%; Score 313; DB 2; Length 400;  
Best Local Similarity 26.8%; Pred. No. 2.4e-17;  
Matches 115; Conservative 57; Mismatches 171; Indels 86; Gaps 14;

QY 26 FDMATNP--SNLSGVQDA-----W-----AVIQESNPEDLV-----WTRCNG 60  
Db 12 FDTDPFASITSDVEHAREKSWATTPYGLAVLYEQVNRLLKHKRLROGSAAMPANH 71  
QY 61 -----GHWATRGQLIREAYEDYRHFSSSECPIPRRAGEAYDFIPSMDBPEOROPRAL 114  
Db 72 VTGEPFAEMFAS-----WLNKEGE-----EHHRLRL 99  
QY 115 ANOVGMVVDKLENRIOELACSLIESLRPOGOCNFTEDYAEPPRIEMLLAGPREDI 174  
Db 100 MNPAFSPKLIGSLVPRFOALANELLDFAPRDCFEVSEFAEPAARVIAIMGLPREE- 158  
QY 175 PHLKYLTDQ-----MTRPDGSMFAEAKELDYDLPILIEQRQKGTDAISVAN 225  
Db 159 --WKVISTESATIGLAVTLREDELPRIEAAVORLYEYSEDLADRANPRDPMITLVN 216  
QY 226 GOVNGRP-----ITSDEAKRMFGLLVGGIDTVVNFLSFSMEFLAKSPENHOELIERPERI 281  
Db 217 A-----SRPDDGRLSKELRDALILLIFGFPDTRQOLAMQTKHKNHDDKRLAGERDILG 273  
QY 282 PANCEELLRRFSIVA-DGRILTSYEFHGVOLKKGDQILLPOMLSGLDERKNACPMHVD 340  
Db 274 GKAAEEVMRVNPTVRWVTRREVDEFEYEGVTLKAGTVNHLSESAGDPR--VEPQFDI 331  
QY 341 SRQKSVHTTGHSHLCLGHLARREIIVTLKEMLTRIPDSIAPGAQIQHKSGIVSGVQ 400  
Db 332 TAERKPHFGFGGVHCLGHFVARSMSSEALPLARLRDPHLEPGATWLPDGS-NTGPN 390  
QY 401 ALPLVMDPA 409  
Db 391 TLRPGTTPA 399

RESULT 15  
Q59831 PRELIMINARY; PRT; 410 AA.  
AC Q59831;  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, last sequence update)  
DE 01-NOV-1999 (TReMBLrel. 12, last annotation update)  
DE CYTOCHROME P450 SCA-2 (EC 1.14.14.1).  
OS Streptomyces carophilus.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SANK 62585;  
RX MEDLINE; 96001248.  
RA WATANABE I., NARA F., SERIZAWA N.;  
RT "Cloning, characterization and expression of the gene encoding  
RT cytochrome P-450sca-2 from Streptomyces carophilus involved in  
RT production of pravastatin, a specific HMG-CoA reductase inhibitor.";  
RL Gene 163:81-85(1995).



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2000, 06:41:30 ; Search time 2798.97 seconds  
(Without alignments)  
591.079 Million cell updates/sec

Title: US-09-246-451-16

Perfect score: 927

Sequence: 1 atgcagtaacctacatc.....gagtgtaacagcaactct 927

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 972840 segs, 892348106 residues

Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_cm:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl1:\*  
10: gb\_pl2:\*  
11: gb\_pl3:\*  
12: gb\_ro:\*  
13: gb\_sts:\*  
14: gb\_sy:\*  
15: gb\_un:\*  
16: em\_fun:\*  
17: em\_hum1:\*  
18: em\_hum2:\*  
19: em\_in:\*  
20: em\_cm:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_sy:\*  
29: em\_un:\*  
30: em\_vl:\*  
31: gb\_hlg1:\*  
32: gb\_hlg2:\*  
33: gb\_hlg3:\*  
34: gb\_hlg4:\*  
35: em\_ba1:\*  
36: em\_ba2:\*  
37: em\_hum3:\*  
38: em\_hum4:\*  
39: gb\_pl4:\*  
40: gb\_hlg3:\*  
41: gb\_hlg4:\*  
42: gb\_hlg5:\*  
43: gb\_hlg6:\*

44: gb\_hlg7:\*  
45: em\_hlg1:\*  
46: em\_hlg2:\*  
47: em\_hlg3:\*  
48: em\_hum5:\*  
49: gb\_pl3:\*  
50: gb\_pl5:\*  
51: gb\_hlg8:\*  
52: gb\_hlg9:\*  
53: gb\_hlg10:\*  
54: gb\_hlg11:\*  
55: gb\_hlg12:\*  
56: gb\_hlg13:\*  
57: gb\_hlg14:\*  
58: gb\_hlg15:\*  
59: gb\_hlg16:\*  
60: gb\_hlg17:\*  
61: gb\_hlg18:\*  
62: em\_hlg4:\*  
63: em\_hlg5:\*  
64: em\_hlg6:\*  
65: em\_hlg7:\*  
66: em\_hum6:\*  
67: gb\_hlg18:\*  
68: gb\_hlg19:\*  
69: gb\_hlg20:\*  
70: gb\_hlg21:\*  
71: gb\_hlg22:\*  
72: gb\_hlg23:\*  
73: gb\_hlg24:\*  
74: gb\_hlg25:\*  
75: gb\_hlg26:\*  
76: gb\_hlg27:\*  
77: gb\_hlg28:\*  
78: gb\_hlg29:\*  
79: gb\_hlg30:\*  
80: gb\_hlg31:\*  
81: gb\_vil1:\*  
82: gb\_vil2:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	925.4	99.8	955	5 A00740	A00740 A.rustlicana
2	925.4	99.8	955	5 A00741	A00741 A.rustlicana
3	925.4	99.8	955	14 SYNHRPER	J05552 Synthetic h
4	843.8	91.0	1072	5 A30294	A30294 Synthetic I
5	843.8	91.0	1072	5 AR000440	AR000440 Sequence
6	843.8	91.0	1072	5 AR068580	AR068580 Sequence
7	564.6	60.9	927	5 A06250	A06250 Synthetic H
8	524	56.5	924	5 E01651	E01651 CDNA encodi
9	493.6	53.2	1236	7 ATPXCB	X71794 A.thaliana
10	490.4	52.9	1229	7 HRAHRPCC	M60729 A.rustlicana
11	481.4	51.9	1195	7 ATPEROX5	X98777 A.thaliana
12	481.4	51.9	1195	7 ATPXR3GE	X98315 A.thaliana
13	304.4	32.8	1202	7 PTPXP2PER	X97349 P.trichocar
14	301.8	32.6	1143	7 POPP01	D83224 Populus nig
15	297.2	32.1	1371	7 PTPXP4PER	X97351 P.trichocar
16	296.4	32.0	1461	7 PTPXP1PER	X97348 P.trichocar
17	295.4	31.9	1222	7 PTPXP3PER	X97350 P.trichocar
18	292	31.5	1115	7 POPPA	D30652 Populus kit
19	290.4	31.3	1268	8 LINUMROX	L07554 Linum usita
20	275.4	29.7	1170	7 POPPB	D30653 Populus kit
21	257.2	27.7	2912	7 HRAHRPCA	M37156 A.rustlicana
22	250.4	27.0	1176	8 AF149277	AF149277 Phaseolus
23	244.4	26.4	3289	7 HRAHRPCB	M37157 A.rustlicana
24	237.6	25.6	85109	8 ATT2J13	AL132967 Arabidops

25	237.6	25.6	129757	8	ATP2K15	AL132956	Arabidops
26	237.2	25.5	3919	7	ATHL3036	AJ133036	Arabidops
27	236.4	25.6	3436	7	ATHRXCA	M58380	Arabidopsis
28	217	23.4	1169	8	ALFPXDC	L36157	Medicago sa
29	216.8	23.4	1220	7	MSRNAPFLA	X90692	M. sativa mr
30	216	23.3	1212	7	ATP2TPA2A	X99992	A. thaliana
31	213.2	22.7	1287	8	AF007211	AF007211	Glycine m
32	207	22.3	3095	7	HRAPRXC2	D90115	Horse radish
33	204	22.0	1311	8	SSNPEROXIB	L36111	Stylosanthe
34	201.4	21.7	2459	8	AF149280	AF149280	Phaseolus
35	199.2	21.5	1238	7	MSRNAPFLB	X90693	M. sativa mr
36	195	21.0	1287	7	MSRNAPFLC	X90694	M. sativa mr
37	192.6	20.8	104386	8	AT132A17	AL161813	Arabidops
38	192.6	20.8	172675	41	AC007635	AC007635	Arabidops
39	192.6	20.8	199362	8	ATCHRIV24	AL161512	Arabidops
40	189.4	20.4	104386	8	AT132A17	AL161813	Arabidops
41	189.4	20.4	172675	41	AC007635	AC007635	Arabidops
42	189.4	20.4	199362	8	ATCHRIV24	AL161512	Arabidops
43	182.8	19.7	1467	8	PAB250121	AJ250121	Picea ab1
44	180	19.4	1320	49	AF244923	AF244923	Spinacia
45	178.4	19.2	1271	8	ALFPXDA	L36156	Medicago sa

ALIGNMENTS

RESULT 1  
LOCUS A00740 955 bp DNA PAT 28-JAN-1993  
DEFINITION A.rusticana synthetic gene for peroxidase.  
ACCESSION A00740.1 GI:14602  
VERSION A00740.1  
KEYWORDS  
SOURCE  
ORGANISM  
horseradish.  
Armoreria rusticana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
Rosidae; Caprales; Brassicaceae; Armoreria.  
1 (bases 1 to 955)  
SYNTHETIC GENE  
AUTHORS  
TITLES  
JOURNAL  
Patent: WO 8901424-A 2 20-APR-1989;  
British Bio-Technology Ltd  
Location/Qualifiers  
1..955  
/organism="Armoreria rusticana"  
/db\_xref="taxon:3704"  
11..940  
/note="synthetic gene"  
/codon\_start=1  
/product="Horseradish peroxidase"  
/protein\_id="CA00083.1"  
/db\_xref="GI:14603"  
/translation="MQLPTFYDNCSPNPNIVROTIYNELNSDPRIASILRLHFHD  
CFVNGCDASILLDNFTSFPERKDAFGNANASAGFVIDRMKAIVESACRYTSCADLT  
T1AAOOSVTLAGPSWVRPLGRDLSLOALDLANMLPAPFTLLPOLKDSFYNGLNR  
SSDILVLSGHTFGKNOCFRIAMDRLYNFENGTGLPDLTNTYLOTIRGLCPLNGNLSA  
LVDFDLRPTTFIDNKRYVNLFEOKGLIOSDOELFSSPNATDITPLVRSRANSTQTFEN  
AFVEAMDRCNITPLTGTGOTRINCRVNSNS"

BASE COUNT 246 a 269 c 221 g 219 t  
ORIGIN

Query Match 99.8%; Score 925.4; DB 5; Length 955;  
Best Local Similarity 99.9%; Pred. No. 3.3e-270;  
Matches 926; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atcagatcaaccctacatclacagcaatagctgtcccaagctgtccaatcgcttgcg 60  
|||||  
Db 11 ATCAGATTAACCCCTACATTCCTACGACAAATAGCTGTCCCAAGCTGTCCACATTCGTTCCG 70  
|||||  
QY 61 gacacaatcgctcaagagctcagatccgatcccaagagctgccttcaatattacgtctg 120  
|||||

Db	71	GACAAATGCTCAACGAGCTCAGATCCGATCCAGATCCGCTGCTTCATATTACGTCG 130
QY	121	cacttcacatcagcttcgtgaatggtgtgcagcagctagcatattactgcaacaccacc 180
Db	131	CACCTTCATACCTCCTTCGTAATGTTGGCAGCGTACGATATTACTGGCAACACCACC 190
QY	181	agttccgcaactgaaaagatgcatlccggaacgctaacacgcgcagggcttccagtg 240
Db	191	AGTTCCGCACTGAAAAGATGATTCATTCGGAGACGCTAAACGCCACGAGGGCTTTCAGTG 250
QY	241	atgatcagatgaagaagctgcgtgtgagtcagcatgcaccaagaacagtcagtgtcgaac 300
Db	251	ATGATCGCATGAAGGCTGCCGTTGATGATGATGCCACAGCAAGTCAATGTTGACAGAC 310
QY	301	ctctgacatcagctgtgcagcaagagcgtgtacactctgcagcagggagccgtctggaagt 360
Db	311	CTGCTGACTATTAGCTGCGCAACAGAGCGTACTCTTGACGCGGACCGCTCTTGAGAGTG 370
QY	361	ccgctcgttcagcgtgactcctccacaggaattcctagatctggccaagcgaacttgct 420
Db	371	CCGCTCGTCGACGTGACTCCCTACAGAGCATTCCTAGATCTGGCCAAACGCAACTTGCC 430
QY	421	gtcccatcttcaacctgcctccagcagctgaagagatagctttagaaacgttggctgaatgc 480
Db	431	GCTCCATTTCTTCAACCCCTGCCACAGCTGAAGATAGCTTTGAAAACGTGGCTGAATCGC 490
QY	481	tcgagtagcctgtgtgcctgtcccgagagacacacatttggaaagaacagtgtaagtc 540
Db	491	TGCAAGTACCTTGTGGCTCTCTCCGAGAGCACACATTTGGAAAGAACAGTGTAGTTTC 550
QY	541	atcatgatalagctctlaaaattcaggaacacttgggttacttcgcccacgcgtgaacact 600
Db	551	ATCATGATAGGCTTACAAATTTCAGCAACACTGGGTTACTCTGACCCACGCTGAAACACT 610
QY	601	acgtatctcagaacactggaagagctgtgtcccaactgaatgcaactcagtgtaactagt 660
Db	611	ACGTATCTCCAGACACTGAGAAGGCTGTGCCACTGATGATGGAACCTCATGACGACAGTG 670
QY	661	gacttgcacctgcggagcccaacacatcttcgataaagaagtaactgtgaatctagaaggag 720
Db	671	GACTTTGATGTGGGGAGACCCACCATCTTGATTAACAAGTACTATGTGAATTCAGAGGAG 730
QY	721	cagaagaagcctgatalacagagtgatcaagaactgtttagcagtlccagaagccactgacac 780
Db	731	CAGAAAGGCTGTATACAGAGTATCAAGACTGTTTAGAGTCCAAAGCCGACCTGACACAC 790
QY	781	atcccaactgttgaagaagtttgccttaactcattaccaaacctctttaaagcctctggaa 840
Db	791	ATCCCACTGGTGAAGATTGTTGCTTACTTACATCAAAACCTTCTTTAAACGCTTCGTGGAA 850
QY	841	gcatgagccgtatggttaacatctaccctctgaagcggtacccaagggcagatctgtc 900
Db	851	GCCATGAGCCGTTATGGGTAACTATTACCCCTCTGTACGGGTACCAAGGCCAGATTCCGTCG 910
QY	901	aactcagagtggtlcaacagcaactct 927
Db	911	AACTCGAGAGTGTCAACAGCAACTCT 937

RESULT 2  
LOCUS A00741/c 955 bp DNA PAT 02-FEB-1993  
DEFINITION A.rusticana synthetic gene (reverse complement) for peroxidase.  
ACCESSION A00741 A00741.1 GI:14604  
VERSION A00741.1  
KEYWORDS  
SOURCE  
ORGANISM  
horseradish.  
Armoreria rusticana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
Rosidae; Caprales; Brassicaceae; Armoreria.  
1 (bases 1 to 955)  
AUTHORS  
Edwards, R.



TITLE	SYNTHETIC GENE									
JOURNAL	Patent: WO 8903424-A 3 20-APR-1989; British Bio-Technology Ltd									
COMMENT	#A00741 is the reverse complement of #A00740.									
FEATURES	location/Qualifiers									
source	1..955									
	/organism="Armoracia rusticana" /db_xref="taxon:3704"									
BASE COUNT	219	a	221	c	269	g	246	t		
ORIGIN										
Query Match	99.8% Score 925.4; DB 5; Length 955;									
Best Local Similarity	99.9% Pred. No. 3.3e-270;									
Mates 926; Conservative	0;	Mismatches	1;	Indels	0;	Gaps	0;			
Oy	1	atgcagtaaacccccaattcttagagaaatgctgtcccaagtggtccaaatcgttc	60							
Db	945	atgcagttAACCCCTACACTTCTACGACAAATAGCTCTCCAAAGTCTCCAACTTCCTGCG	886							
Oy	61	gacacaatcgttcaaaagagctcaagatccagatcccaagatcgtcttcaataltaagtcg	120							
Db	885	GACACAATCGTCAAGAGACTGAGATCCGATCCGAGAGATGCGTCTCAATATATAGTCTG	826							
Oy	121	caattccatgactgctctctggaatggttgcgagcgttagcatactatctgcaacaacac	180							
Db	825	CACCTTCCATGTAGCTGTTGCTGTAATGTTGCGACGCTACATATATCTGGACAAACACC	766							
Oy	181	agttccgacatcgaaaagatgatcttgaggaaagcttaacagcgccagggccttccagt	240							
Db	765	AGTTCCCGACACTGAAAAGATGATCTCGGAAGCGTAAACGCCGCAAGGGCTTCCAGTG	706							
Oy	241	atcgaatcgatbaaagcgctgcgtttagatcaagatgccccagaagatcaatgltgacag	300							
Db	705	ATCGATGCCATGAAGAGCGTCGCTTAGTGAATGAGATGCCACAGACAGTCAAGTGTCAAG	646							
Oy	301	ctgtctgactatagctctgcgcaacagagcgttgactcttgcaagcgcgactcctgtagag	360							
Db	645	CTGCTGACTATATAGCTGCGCCACAGACAGACCGCTGACTTTCGAGGGCGACCGCTCGGAGAGTG	586							
Oy	361	ccgctccggtgcgaagctgaaatcccttaagggatctccagatcttgagccaagccaactgct	420							
Db	585	CCGCTCGGTCGACGTAACCTCTACAGGCATTTCTAGATCTGGCCAAAGCCCAACTTTCCT	526							
Oy	421	gctccatctcttaacctgtccccagcttgaaagatagctttaagaagctggtgtgaatcgc	480							
Db	525	GCTCATTTCTTACCCCTGCCCGCAGTGAAGATTAACCTTAGAAAGCTGGGTGTGAATGCG	466							
Oy	481	tcgagtgaaactgtgtgctctgtccggagagacacacatttggaaagaacagtgtagtgc	540							
Db	465	TCGAGTGAACCTTGTGGCTGTCTGTCGCGGAGGACACATTTGGAAAGAACAGTGTAGTTTC	406							
Oy	541	atcataggaatgagcttataaatltcagcaaacactggtgttaactgtgaecccaagctgaac	600							
Db	405	ATCATGGATAGGCTCTTACAAATTTACGAAACATGTGGTTACCTGACCCCAAGCTGAACCT	346							
Oy	601	acgatctccaaagaaactgagagcctgtgtgccccaaatgaaatlggaaccttaagtgaatg	660							
Db	345	ACGTAATCTCCAGACACTGAGAGGCTTGTGCCCACTGAATGGCAACCTTAAGTCACTAGTG	286							
Oy	661	gactttagatctgcgaaccccaaccaactctctcgataaacaagtactaactgtaalttagagg	720							
Db	285	GACTTTGATCTGCGGAGCCCAACCATCTTCGATTAACAAGTACTATGTGATCTAGAGGAG	226							
Oy	721	cagaaaagccctgataccaagatgataaagaacgtgttagcaagtccagaagccaactgaacc	780							
Db	225	CAGAAAGCCCTGATTCACAGATGATAAAGAACTGTTTACAGATGCACAAAGCCACTACAC	166							
Oy	781	atcccacatggatagaagatttgctaaacttaactaaacttctttaacgaccttgtagaa	840							
Db	165	ATCCCACTAGTGTAGAGATTTTCTACTACTACTCAAACTCTTTTAAGCCCTTCTGTGGAA	106							
Oy	841	gccatcgagccgtatggtgtaacataccctccttgaaagcggtatcccaagagccagatctgctg	900							

DB	105	GCCTAGGACCGCATGCGGTAACCTTACCCCTCTGACGGGATACCCAGGCCAGATTTCGCTG	46
QY	901	aactgcagatgtgtcaacagcaactct 927	
DB	45	AACCTGACAGTGTCTACACAGCAACTCT 19	
RESULT	3		
SYNHRP			
LOCUS			
DEFINITION	SYNHRP	955 bp DNA	SVN 27-Apr-1993
ACCESSION	Synthetic horseradish peroxidase isoenzyme C (HRP-C) subunit		
VERSION	alpha-1 DNA, complete cds.		
KEYWORDS	J05552.1 GI:208493		
SOURCE	peroxidase.		
ORGANISM	Synthetic DNA.		
REFERENCE	Synthetic construct		
AUTHORS	artificial sequence.		
	1 (bases 1 to 955)		
	Smith,A.T., Santama,N., Dacey,S., Edwards,M., Bray,R.C.,		
	Thornely,R.N.F. and Burke,J.F.		
	Expression of a synthetic gene for horseradish peroxidase C in		
	Escherichia coli and folding and activation of the recombinant		
	enzyme with Ca2+ and heme		
	J. Biol. Chem. 265, 13335-13343 (1990)		
JOURNAL	90330683		
MEDLINE	Draft entry and computer-readable sequence for [J. Biol. Chem.		
COMMENT	(1990) In press] kindly submitted		
	by A.T.Smith, 13-JUN-1990.		
FEATURES	location/Qualifiers		
source	1..955		
	/organism="synthetic construct"		
	/db_xref="taxon:32630"		
	11..940		
CDS	/note="synthetic horseradish peroxidase isoenzyme C		
	(HRP-C) subunit alpha-1 (E.C. 1.11.1.7)"		
	/codon_start=1		
	/transl_table=1		
	/protein_id="AA072223.1"		
	/db_xref="GI:208494"		
	/translation="MQLPTPTFDNSCPNVSNIYDITVNELRSDPRIASILRLPHD		
	CEVNGDASILLDDWTSFRTKEKDAFGNANSRGPVIDMRKAAYESACPRVSCADLL		
	TIAAQSVTLAGGSPMSRWPLGRDLSLAFLDIANANPAFFTLPLKDSFERNGLNR		
	SSDIYLVSGHTEFGKNOCRFLMDRLVNSNTGDDPLNTLYTLTLGRLDPLNGLSA		
	LVDDLRPTIEFDNKKYVNL EOKGILIOSOELESSPNADTIPLVASFANSYQTFPN		
	AFVAMBRMGITPLTGTGQIRLNCRVNSN"		
BASE COUNT	246 a 269 c 221 g 219 t		
ORIGIN	1 bp upstream of HindIII site.		
Query Match	99.8% Score 925.4; DB 14; Length 955;		
Best Local Similarity	99.9%; Pred. No. 3.3e-270;		
Matches 926; Conservative	0; Mismatches 1; Indels 0; Gaps 0;		
QY	1	atggaagtaacccacacattctagaacaatagctgtcccaagtggtccaaatcgttcgc 60	
DB	11	ATGCGATTAAACCCCTACATTCTACGACATAAGCTTCCCAACGTCTCCAACTCGTTCG 70	
QY	61	gacacaatcgtcaagaagctcagatccgatcccaagatcgtgtcttcaataataagtcgtc 120	
DB	71	GACACAATCGTCAAGAGACTGAGATCCGATCCGACGATCGCTTCAATATTACGCTG 130	
QY	121	caattccatgaatgcttctgtgaatggtgtgcgaagcttaacataatctctggaacaacacc 180	
DB	131	CACCTTCATGATGCTGTTCGTGTAATGGTTGCGACGCTACATATTACTGGACAAACACACC 190	
QY	181	agttccgaacgaaagaatgatcattcgggaaagcttaacagcgccaggaggtcttcacagt 240	
DB	191	AGTTCCGACGACGAAAGAGATCATTCGGGAACGTTAACAGCGCCAGGGGCTTTTCAATG 250	
QY	241	atgatcgcataaaggtctgcgttgatgcagatgcccacgaacagtcagttgtgtcagac 300	

```
Db 251 ATGATGCGATGAAAGGCTGCCGTTGAGTCAGATGCCCCAGAACAGTCAAGTTGTGCAGAC 310
Oy ctgtctactaatactgacgacagagcggtactcttgcaagcgagaccgtctctgagagtg 360
Db 311 CTGCTGACTATACCTGCGCAACAGACGCTACTCTTGCAAGCGGACCGCTCTTGAGAGTG 370
Oy 361 ccgctcgctgcagcgtactccctactcaagcalttctagatctggccaagcccaacttgct 420
Db 371 CCGCTCGGTGACGTACTGCTCCCTACAGAGCATGCTAGATCTGGCCAAACGCCAATTGCTT 430
Oy 421 gctccattcttcaacctgcccagctggaagatagctttagaacgtgggtctgtaatgc 480
Db 431 GCTCCATTTCTTCACCCCTGCCAGCTGAAGATAGCTTTGAAAACGTGGGTCTGAATCGC 490
Oy 481 tcgaagtgaccttggtgctctgctccgagagacacatttgaaagaacccagtgtaagttc 540
Db 491 TCGAGTGACCTTTGGCTCTGCTCCGAGAGACACATTTGGAAAAGAACCGTGTAGGTTTC 550
Oy 541 atcatgataagcgtctactaaatttcaagcaaacactgggttacttgaccaccaagctgaacact 600
Db 551 ATCATGATATAGGCTGTACATTTTCAGCAACACATGGGTTACTTGACCCCAAGCTGAGACACT 610
Oy 601 acgtatctccagacactgagaggtctgtgccacactgaatggcaacctgaatgactagtg 660
Db 611 ACCTATCTCCAGACACTGAAAGGCTTGTCGCCACTGATGCAACCTCACTGACTAGTG 670
Oy 661 gaattgatactggagaccccaaccatcttcgataaagaactatgtgaattctagaagag 720
Db 671 GACTTTGATCTGCGGACCCCAACCATCTTCGATTAACAAGTACTATGTGATCTAGAGAG 730
Oy 721 cagaagaagcctgaatacagaatgaataagaactgtttagcagtlccagaagccactgaacac 780
Db 731 CAGAAAGGCTGTATACAGAGTGAATCAAGAACTGTTTAGCAGTCCAAAGCCACACGACACC 790
Oy 781 atcccaactgttggaagatgttctgaactctaccctcaaacctctttaaagcctctgagaa 840
Db 791 ATCCCACTGGTGAAGAGTTTGTCTTACTTACTTAACTTAAACCTTTTAAACGCTTCGTG 850
Oy 841 gccatgaccgtatggttaacatctacccctctgaaggtacccaagccagatctgctg 900
Db 851 GCCATGGACCGTATGGGTAACTTACCCTCTTACCGGTACCGGATCCCAAGGCCAGATTCTG 910
Oy 901 aactgcagatggttcaacagcaactct 927
Db 911 AACTGCAGAGTGTCAACAGCACTCT 937

RESULT 4
A30294 1072 bp DNA PAT 09-OCT-1995
LOCUS A30294 Synthetic IGF-II/HRP fusion protein gene.
DEFINITION A30294
ACCESSION A30294
VERSION A30294.1 GI:1247537
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 1072)
AUTHORS IGF-II ANALOGUES
TITLE Patent: WO 9303152-A 9 18-FEB-1993;
JOURNAL Location/Qualifiers
FEATURES
source 1..1072
/gene="IGF-II"
/db_xref="taxon:9606"
misc_feature 11..856
/gene="HRP"
misc_feature 11..856
/gene="HRP"
CDS 11..1060
/codon_start=1
```

```
/product="IGF-II/HRP fusion protein"
/protein_id="CAA02021.1"
/db_xref="GI:1247538"
/translation="MQLPPPEADNSQPNVSNIVRDIYVELRSDPRIASILRLRHHF
CFVNGCDASILLDNNTTSFPEKDAFGNANASAGFPVIDMKAAVESACRIVSCADL
TIAQOSTVTLAGSPSWRPLGRDRSDLOALELDLANANLPAPEFTLLPOLKSPFNGLNR
SSDLVALSGHFNKNOCRFIMDRLVNENSCIGLPPPTLNTYGLQTLRGICPLNGINLSA
LVDFDLRPTTIFEDNKRYVNLFEOKGLIOSDELSSPNATDRIPIVRSFANSSTGFEN
AFVEMAYRPSFTLCGGELVDLTQLFVCGDRGRFYFSPRARSRSRGIVIEECFNSCD
LALLETYCATPAPKSE"
misc_feature 857..1057
/partial
/gene="IGF-II"
/gene="IGF-II"
BASE COUNT 259 a 304 c 255 g 254 t
ORIGIN
Query Match 91.0%; Score 843.8; DB 5; Length 1072;
Best Local Similarity 99.8%; Pred. No. 2.1e-245;
Matches 845; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 1 atgcagttaacccctacatctactacgaacatagctgtcccaagctgtcccaacatcgttcgc 60
Db 11 ATCAGATTAAACCCCTACATCTACGACAAATAGCTGTGCCAACGTGCCAACATCGTTGCG 70
Oy 61 gacacaatgctaaagagctcagatccgataccagatgcgtctccaataataagctcg 120
Db 71 GACCAATCGTCAACGAGCTCAAGTCCGATCCGAGATCCCTCTTCAATATACGCTTG 130
Oy 121 cacttcacatgactgcttgtaatggtlttgagacgtactacatctacttggaacacacacc 180
Db 131 CACTTCCATGACTGCTTTCGTAATGCTGCGAGCTTACGCTATTTACTGGAACACACCACC 190
Oy 181 aglttcgcactgaaaaagatgcatctgcggaacgcctaacagcgccagggcttccagtg 240
Db 191 AGTTTCGCGACTGAAGAGATGATTCGGGAAACGCTAACAGCCGAGGGGCTTCCAGTG 250
Oy 241 atcgatcgcataagagctgcgcttgtagtcaagcatgcccagaacagtcagttgcaagac 300
Db 251 ATGATGCGCATGAAGGCTGCGGTTGAGTCAGATGCCACAGCAAGTCAATGTTGAGAC 310
Oy 301 ctgtctactaatactgacgacagagcggtactcttgcaagcgagaccgtctctgagagtg 360
Db 311 CTGCTGACTATACCTGCGCAACAGACGCTACTCTTGCAAGCGGACCGCTCTTGAGAGTG 370
Oy 361 ccgctcgctgcagcgtactccctactcaagcalttctagatctggccaagcccaacttgct 420
Db 371 CCGCTCGGTGACGTACTGCTCCCTACAGAGCATGCTAGATCTGGCCAAACGCCAATTGCTT 430
Oy 421 gctccattcttcaacctgcccagctggaagatagctttagaacgtgggtctgtaatgc 480
Db 431 GCTCCATTTCTTCACCCCTGCCAGCTGAAGATAGCTTTGAAAACGTGGGTCTGAATCGC 490
Oy 481 tcgaagtgaccttggtgctctgctccgagagacacatttgaaagaacccagtgtaagttc 540
Db 491 TCGAGTGACCTTTGGCTCTGCTCCGAGAGACACATTTGGAAAAGAACCGTGTAGGTTTC 550
Oy 541 atcatgataagcgtctactaaatttcaagcaaacactgggttacttgaccaccaagctgaacact 600
Db 551 ATCATGATATAGGCTGTACATTTTCAGCAACACATGGGTTACTTGACCCCAAGCTGAGACACT 610
Oy 601 acgtatctccagacactgagaggtctgtgccacactgaatggcaacctgaatgactagtg 660
Db 611 ACCTATCTCCAGACACTGAAAGGCTTGTCGCCACTGATGCAACCTCACTGACTAGTG 670
Oy 661 gaattgatactggagaccccaaccatcttcgataaagaactatgtgaattctagaagag 720
Db 671 GACTTTGATCTGCGGACCCCAACCATCTTCGATTAACAAGTACTATGTGATCTAGAGAG 730
Oy 721 cagaagaagcctgaatacagaatgaataagaactgtttagcagtlccagaagccactgaacac 780
```

Db 731 CAGAAAGCCCTGATACAGAGTATCAAGACTGTTAGCAGTCCAAAGCCCACTGACACC 790  
QY 781 atccacatggtgagaagtttctgtaacttaactcaaccttcttaacgcttcgtggaa 840  
|||||  
Db 791 ATCCCACTGGTGAAGAGTTTGGTAACTACTACCAACTCTTTAAACGCTTGTGGAA 850  
QY 841 gccatgg 847  
|||||  
Db 851 GCTATGG 857

RESULT 5  
AR000440  
LOCUS AR000440 1072 bp DNA PAT 04-DEC-1998  
DEFINITION Sequence 9 from patent US 5736363.  
ACCESSION AR000440  
VERSION AR000440.1 GI:3962971  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1072)  
AUTHORS Edwards, R. Mark and Bawden, L.  
TITLE IGF-II analogues  
JOURNAL Patent: US 5736363-A 9 07-APR-1998;  
FEATURES Location/Qualifiers  
source 1..1072  
BASE COUNT 259 a 304 c 255 g 254 t  
ORIGIN

Query Match 91.0%: Score 843.8; DB 5; Length 1072;

Best Local Similarity 99.8%: Pred. No. 2,1e-245;

Matches 845; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atgcagttaacccttaactcttaagcaatagctgtcccaagctgtccaacatcgttcgc 60  
|||||  
Db 11 ATGCAGTTAACCCCTCATATTCTAGCAATAGCTGTCCCAAGCTGTCAACATCGTTCCG 70  
QY 61 gaaacaatcgtcaagagctcagaatccgataccagagatcgctgtccaatataagctcg 120  
|||||  
Db 71 GACACAATGCTCAACAGAGCTCAGATCCGATCCAGATCGCTTCAATATTACGCTTG 130  
QY 121 caattccagtaactgctctgtaagtgttcgagcgttagcatattacttgagcaaacacc 180  
|||||  
Db 131 CACTTCATGACAGCTCTTCTGTAATGTTGCGACGCTAGCATATTACTGGACACACCACC 190  
QY 181 agttccgactgaaagagatgcatcgcggaacgctaacagcgccaggggcttccagt 240  
|||||  
Db 191 AGTTCCGCACTGAAGAGATGATTCGGGAACGCTAACAGCGCGCTTTCAGTG 250  
QY 241 atcgatcgaatgaagcgtcgctgtgagtcagcatgcccacgaacatcagttgtcaag 300  
|||||  
Db 251 ATCGATCGATGAAGGCTCCCTGTTGAGTCAGCATGCGCCACAGAACATGTCGAGAC 310  
QY 301 ctgctactactaactcgtcgcaacagagcgtgaactcttcgagcgagaccgttcgtgaga 360  
|||||  
Db 311 CTGCTACTATTAGCTCGCAACAGAGCGTACTCTTGGACGGGACGCTCTGGAAGTG 370  
QY 361 ccgctcgtcgaactgactccctactcaagcatctcctagatctgccaacgcaacttgct 420  
|||||  
Db 371 CGCGTGGTGCAGAGTACCTCCCTACAGGCAATCTCTAGATCTGGCCAACGCAACTTGCT 430  
QY 421 gctccattcttcaacctgcccacgctggaaggaatagctttagaagcgtgggtctggaatcgc 480  
|||||  
Db 431 GCTTCATTTCTTCCACCTGCCCCAGCTGAAGGATAGCTTTAGAAACGTGGGCTCGAATCGC 490  
QY 481 tcgaagtgaacttggctctgttcgcgagagacacacatttgaaagaacagtgtaagtttc 540  
|||||  
Db 491 TCGAAGTGACTTTGTGGCTGTGTCGGAAGACACACATTTGGAAAGAACAGTGTAGTTTC 550  
QY 541 atcatgatagcgtctacaatttcagaaacactgggttactctgaccaccaagctgaacact 600

Db 551 ATCATGATAGGCTCTACAAATTTACAGCAACACTGGGTTACCTGACCCCAAGCTGAACACT 610  
QY 601 acgatactcagaacactgtgagaagctgtgtgccactgaatggcaactcagtgtaactagtg 660  
|||||  
Db 611 ACCTATCTCCAGACACTGAAGAGCTTGTGCCCACTGAATGGCAACCTCACTGACATAGTG 670  
QY 661 gaattgatactgcgagccccaacatcttcgaataacgaactatgtgaattcagaaggag 720  
|||||  
Db 671 GACTTTGATGTGGGAGCCCAACCATCTTCTGATTAACAAGTACTATGTAATCTAGAGAG 730  
QY 721 cagaagggcctgatacagagtgatacaagaactgtttagcagtlccagaagcgaactgacac 780  
|||||  
Db 731 CAGAAAGCCCTGATACAGAGTATCAAGACTGTTAGCAGTCAAAAGCCCACTGACACC 790  
QY 781 atccacatggtgagaagtttctgtaacttaacttaactcaaaccttctttagcgttcgtggaa 840  
|||||  
Db 791 ATCCCACTGGTGAAGAGTTTGGTAACTACTACCTTACCAACTTCTTTAAACGCTTGTGGAA 850  
QY 841 gccatgg 847  
|||||  
Db 851 GCTATGG 857

RESULT 6  
AR068580  
LOCUS AR068580 1072 bp DNA PAT 29-SEP-1999  
DEFINITION Sequence 9 from patent US 5854025.  
ACCESSION AR068580  
VERSION AR068580.1 GI:6000787  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1072)  
AUTHORS Edwards, R. Mark and Bawden, L.  
TITLE IGF-II analogues  
JOURNAL Patent: US 5854025-A 9 29-DEC-1998;  
FEATURES Location/Qualifiers  
source 1..1072  
BASE COUNT 259 a 304 c 255 g 254 t  
ORIGIN

Query Match 91.0%: Score 843.8; DB 5; Length 1072;

Best Local Similarity 99.8%: Pred. No. 2,1e-245;

Matches 845; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atgcagttaacccttaactcttaagcaatagctgtcccaagctgtccaacatcgttcgc 60  
|||||  
Db 11 ATGCAGTTAACCCCTCATATTCTAGCAATAGCTGTCCCAAGCTGTCAACATCGTTCCG 70  
QY 61 gaaacaatcgtcaagagctcagaatccgataccagagatcgctgtccaatataagctcg 120  
|||||  
Db 71 GACACAATGCTCAACAGAGCTCAGATCCGATCCAGATCGCTGCTTCAATATTACGCTTG 130  
QY 121 caattccagtaactgctctgtaagtgttcgagcgttagcatattacttgagcaaacacc 180  
|||||  
Db 131 CACTTCATGACAGCTCTTCTGTAATGTTGCGACGCTAGCATATTACTGGACACACCACC 190  
QY 181 agttccgactgaaagagatgcatcgcggaacgctaacagcgccaggggcttccagt 240  
|||||  
Db 191 AGTTCCGCACTGAAGAGATGATTCGGGAACGCTAACAGCGCGCTTTCAGTG 250  
QY 241 atcgatcgaatgaagcgtcgctgtgagtcagcatgcccacgaacagtgtaagtttc 300  
|||||  
Db 251 ATCGATCGATGAAGGCTCCCTGTTGAGTCAAGTCCCAAGCAACATGTCGAGAC 310  
QY 301 ctgctactactaactcgtcgcaacagagcgtgaactcttcgagcgagaccgttcgtgaga 360  
|||||  
Db 311 CTGCTACTATTAGCTCGCAACAGAGAGTGACTCTTTCAGCGGACGCTCTGGAAGTG 370

QY 361 ccgctcgctcgacgctgactccctacagcaatccctagatctgagccaaagcgaattgct 420  
|||||  
Db 371 CCGCTGGTGGAGAGTACTCCCTACAGGCAATTCCTAGATCTGGCCAAAGCCAACTTGCCCT 430  
QY 422 gctccattctccaccctgcccagcgaagagtagctttagaagcgtggctgaaatcgc 480  
|||||  
Db 431 GCTCCATTCTCACCCTCGCCAGCTGAAGGATAGCTTTAAGAAACGTGGGTGAAATCCG 490  
QY 481 tcgaagtacattgagctctgctccgagagacacacatttgaaagaaacagtgtaagtc 540  
|||||  
Db 491 TCGAGTACCTTGTGGCTCTGTCGGAGGACACACATTTGGAAAGAACAGTGTAGGTTTC 550  
QY 541 atcaatgataagctctacaattcagcaaacactgggttaacctgaagcccaagctgaacact 600  
|||||  
Db 551 ATCATGATAGGCTCTACATTTACAGCAACACACTGGGTTACTGACCCCAAGCTGAACACT 610  
QY 601 agctatctccagacacttgagagctgtgtcccaactgaatggcaacctcaagctgacagtg 660  
|||||  
Db 611 AGCTATCTCCAGACACTGAGAGGCTTGCCCACTGAATGGCAACCTCAGTGCATAGTTC 670  
QY 661 gactttgactcgcgagcccaacacatctcgataacaagtaactatgtgaaatcagaagag 720  
|||||  
Db 671 GACTTTGATCTGGGGAGCCCAACCATCTTCGATTAACAAGTACTGATGATGAGAGAG 730  
QY 721 cagaaagcctgatacagagtgatacaagaactgttttagcagtcacagccagctgaacacc 780  
|||||  
Db 731 CAGAAAGGCGCTGATACAGAGTGTATCAAGAACTGTTAGCAGTCCAAACGCCCACTGACACC 790  
QY 781 atcccactgtgtagaagtttctgaacttactcaaaccttcttaacgcttcgtgaa 840  
|||||  
Db 791 ATCCCATCTGGTAGAAGTTTGCTAATCTACTCAAACTCTTTAAAGCCCTTCGTGAGAA 850  
QY 841 gccatag 847  
|||||  
Db 851 GCTATGG 857  
RESULT 7  
A06250 927 bp DNA PAT 08-MAY-1996  
LOCUS Synthetic HRP isoenzyme C gene.  
DEFINITION A06250  
ACCESSION A06250.1 GI:1566731  
VERSION A06250.1 GI:1566731  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1 (bases 1 to 927)  
AUTHORS Chiswell,D.J. and Ortlepp,S.A.  
TITLE DNA sequence coding for HRP enzyme  
JOURNAL Patent: EP 0299682-A 19 18-JAN-1989;  
AMERSHAM INTERNATIONAL plc  
FEATURES  
source location/Qualifiers  
1..927  
/Organism="synthetic construct"  
/db\_xref="taxon:32630"  
BASE COUNT 247 a 263 c 198 g 219 t  
ORIGIN  
Query Match 60.9%; Score 564.6; DB 5; Length 927;  
Best Local Similarity 75.7%; Pred. No. 1.5e-160;  
Matches 699; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

Db 121 TTTTCATGACTGCTTTTAAACGGATGTGACGGGTCCATCTCTCTGGAATACAAAC 180  
QY 184 ttccgcagtaaaaggaatgattcctgggaacgctaaagcgaaggcttctcaagtgatc 243  
|||||  
Db 181 TTCCGAACTGAGAAAGACCATTCGGGAACGCCAATTCGGCCCGGGGCTTTCCAGTATC 240  
QY 244 gatcgatgaagctgcgcttgatgataagtcaccaagacagtcagctgtgcaagctg 303  
|||||  
Db 241 GATCGCATGAAGGCTGCAATCCAGTGGCTCGCCGAGAACTGTTTCGTGTGACAGCTG 300  
QY 304 ctgactatagctgcgagacaagagcgctgactctgcaagcggagccgctcctggaagtgccg 363  
|||||  
Db 301 CTGACCATGCGAGCCCGAGCAATCGGCACCTGACCGCGGGCCCTCATGGAGATACCA 360  
QY 364 ctcggtcgagcgtgactcccttaagagatctcctagatctggtccaagccaactgctgtc 423  
|||||  
Db 361 TTAGAGGCTGTGACTCCCTACAGGCTTTCTGATGATCTAGTAAATCAATCTACCGGGC 420  
QY 424 ccattctcacctgcccagctgaaagatagctttagaagcgtgtgtctgtaactcgtc 483  
|||||  
Db 421 CCATTTTACTCTACCAACTTAAGATTCATTTCTGTAAGCTGAGCTTGAACGCTCG 480  
QY 484 agtgcctgtgctctgtcgcggaagacaacaatttgaaagaaacccagtgtagtcatc 543  
|||||  
Db 481 AGTATCTAGTAGCTCTTTCTGTGGGCCACACTTTTGAAGAAACAGTGCAGATTCAATC 540  
QY 544 atggaagctctacaatttcaagcaacactgggttaacctgcaagccagctgaaacctag 603  
|||||  
Db 541 ATGGACCGATTGTACACTTTTCAAACTGACATACCGGATCCACGCTCAATAGAGACC 600  
QY 604 tatctccagacactgaaagactgtgtcccaactgaatgtaaacctcagctgactagtgcac 663  
|||||  
Db 601 TACCTACAGACACTCCGGGACTATGCCCTTTAAAGGAACCTTAAGCCGCTAGTCGAC 660  
QY 664 ttgatctgcggaaccccaacalcttgataaagaactatgtgaatagaaagagcag 723  
|||||  
Db 661 TTTGACCTACGACGCAACAATCTTCGACAACAAGTACTATGTAATCTAGAGAACAA 720  
QY 724 aaaagcctgatacagaatgataagaagactgttagaagtcgaagccagccactgaacccatc 783  
|||||  
Db 721 AAAGGACTTATTCATCTGATCAGAGGCTTCTCTCATCCCAATGCAACAACAGATC 780  
QY 784 ccaactgtgtagaagtttctgaacttactcaaaccttctttaaagcctctgtggaagcc 843  
|||||  
Db 781 CCACGTAGTGCATAGCTTTCGAACTCAACAAACCTTTCTTAATGCATTTCTGTAAGGCGC 840  
QY 844 atggacgtaatggtgaatacattacccctctgacggtgtaaccaagccagaattcgttgaaac 903  
|||||  
Db 841 ATGGATCGAATGGAATATTCACGCGCCTACGCGGTACCAAGTCAGATCAGTGTGAAC 900  
QY 904 tgcagagtggtcaacaagaacac 926  
|||||  
Db 901 TCCCGAGTGTAAACTCGAATCTC 923  
RESULT 8  
E01651 924 bp RNA PAT 29-SEP-1997  
LOCUS E01651  
DEFINITION cDNA encoding horseradish peroxidase.  
ACCESSION E01651  
VERSION E01651.1 GI:2169904  
KEYWORDS JP 1988207386-A/1.  
SOURCE Armoreria rusticana.  
ORGANISM Armoreria rusticana.  
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
Rosidae; Capareales; Brassicaceae; Armoracia.  
AUTHORS Okada,H., Niina,A. and Yamada,Y.  
TITLE STRUCTURAL GENE OF HORSE RADISH PEROXIDASE  
JOURNAL Patent: JP 1988207386-A 1 26-AUG-1988;  
TOYOBO CO LTD

COMMENT	OS	(horseradish)
PN	JP 1988207386-A/1	
PD	26-AUG-1988	
PF	19-FEB-1987 JP 1987037623	
PI	OKADA HIROSUKE, NIINA ATSUSHIKO, YAMADA YASUYUKI	
PC	CI12N15/00, CI12N1/20, CI12N9/08, (CI12N1/20, CI12R1:19), (CI12N9/08, PC	
	CI12R1:19);	
CC	strandedness: Double;	
CC	topology: Linear;	
CC	hypothetical: No;	
CC	anti-sense: No;	
CC	*source: strain=Armoreria lapathifolia, Armoreria rusticana;	
CC	*source: tissue=leaf-root;	
CC	*source: clone=PSK1;	
CC	Feature is identified by similarity;	
FH	Key	Location/Qualifiers
FT	FT	1. .924
FEATURES	source	Location/Qualifiers
		1. .924
		/organism="Armoreria rusticana"
		/db_xref="taxon:3704"
BASE COUNT	256 a	223 c 188 g 257 t
ORIGIN		
Query Match	56.5%;	Score 524; DB 5; Length 924;
Best Local Similarity	72.9%;	Pred. No. 3.2e-148;
Matches 674;	Conservative 0;	Mismatches 250; Indels 0; Gaps 0
4	cagtaaccctcatctctacgacatagctgtcccaacgtgttcccaacatcgtctgcgac	63
1	CAACTTACCCCTTACCTTACGACAAATTCATGTCTTAATGCTCTACATCGACGGAT	60
64	acaatcgttcaagaagatcagatccgatcccaagatcgcgtctcaatatcagttcgac	123
61	ACTATTGTCAAAGACTTAAGATCAACCTCGTATTGGCCGACGACATCCCTGCTTCAC	120
124	ttccatgactgttctgtgaatgtttgcgcagctagacatcttactggagaacaccacagt	183
121	TTCCACGACGCTTGTGTTAATGTTGTGACGCAATGCATGCTTTGTTACACAAACATCA	180
184	ttccgactgaaagaagatgacatctcggaagcgttaacagccagccagggcttccagtgac	243
181	TTTCGAACGAGAAAGATGCTGTTGGAAACGCAAACTCGGCAAGAGATTTCCAGAT	240
244	gacgcacgaagagcgtgcgtgtgagtcacgaatgcccaagcagtcagttgtgcagactg	303
241	GATTGAATCAAAAGCCGCGGTGAGAGTGCATCCCAAGAACCGTTTCATGCCACATTTG	300
304	ctgactatagctgcgcacaagaagcgttgactcttcaaggagaccggtccctggagatgccg	363
301	CTCACCATTTGCAGCTCAAAATCTGTTCATTTGGGGGGAGTCTTCTTGGAGATTCCT	360
364	ctcgcgtcagcgtgactccctacagcagatccctcagatctctgcacacgcacactgtcgt	423
361	TTGGGCGAAGAGATAGCTTACACGATTTCTGGATCTTGCTAATTCAAATCTTCCAGCT	420
424	ccaattcttaaccctgcgccagcttgaagtagctttagaagcttgggtctgaatcgctg	483
421	CCATTCTTCACACTTCCACAACTTAAAGACACTTTAGAAATGTTGGCCTCAACCGTCT	480
484	agtgacatttggtgcctgtgcgcgaggaagacacacatttgaagaagaacacagtgatcacc	543
481	TCTGATCTCTGTGCACTGTCCGGGGGCCACACATTTGGTAAAAAATAGTGTGGTTATT	540
544	atcgatagccttacaattctcagcaaacacatgtgttactcgcgacccacagctgaacactag	603
541	ATGGACAGATTATACAACTTACGAAACACGGGTTTACCGGATCTTACTTCACACTTACT	600
604	tatctccacaacatcgagagcgttgcgccactgaatggacactcagtgcaactagtcactagtgc	663

Db	601	TATCTCCAAACTCTTCGGGACGTATGTGCCCTCAATGTAATCTAAGCGCTTTGGTGAT	660
Oy	664	lttgatctcgagaccaccaacatcttcgataaagaactcgtgaatcgtagaagagca	723
Db	661	TTTTGCATTACGCTACGCCACCATTTTGTGACAACAATACTATGTGAATCTCGAAGAGGAA	720
Oy	724	aaagccgatcacagagatgcgaagaactgitttagcagttccaagacgcccatcgacaacatc	783
Db	721	AAAGAGCTTATCCAAAGGACCCAMAGAAGTTGTCCTTACCCCAATGCCATGCACAATC	780
Oy	784	ccacaggcgagaagtltttgcctaactctactcaaacctctttaagccttcgttgaagcc	843
Db	781	CCTTTGGGAGACATCACTTTTGCTATATAGCACACAAACATTTCTCAATGCAATTTGTGGAGCG	840
Oy	844	atgacccgatcyggytaacattaccacctctgacggtyaacccaagaccagatcgtctgaac	903
Db	841	ATGCTATGATAGTAGGAAACCATTAACCTCTTACAGAGACTCAAGGACAGATCAGGTTGTAAT	900
Oy	904	tgcagagtgcgtcaacagcaactctc 927	
Db	901	TGTAGGCTGTGTAACCTCAACTCT 924	
RESULT	9		
ATPRXCB	LOCUS	APPRXCB 1236 bp mRNA PLN 04-NOV-1994	
DEFINITION	A.thaliana prxcb mRNA for peroxidase.		
ACCESSION	X71794		
VERSION	X71794.1 GI:405610		
KEYWORDS	basic peroxidase; heme; prxcb gene.		
SOURCE	thale cress.		
ORGANISM	Arabidopsis thaliana		
REFERENCE	Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
AUTHORS	euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;		
TITLE	Rosidae; Caprales; Brassicaceae; Arabidopsis.		
JOURNAL MEDLINE	1 (bases 1 to 1236)		
REFERENCE	Intapruk,C., Takano,M. and Shimmyo,A.		
AUTHORS	Nucleotide sequence of a new cDNA for peroxidase from Arabidopsis		
TITLE	thaliana		
JOURNAL	Plant Physiol. 104 (1), 285-286 (1994)		
MEDLINE	94159792		
REFERENCE	2 (bases 1 to 1236)		
AUTHORS	Shimmyo,A.		
TITLE	Direct Submission		
JOURNAL	Submitted (30-APR-1993) A. Shimmyo, Osaka University, Dept of		
COMMENT	Biochemistry, Faculty of Science, Suita, Osaka 565, JAPAN		
FEATURES	Related sequences: Intapruk,C. et al., Gene. 93:237-241 (91) &		
source	Intapruk,C. et al., J. Ferment. Bioeng. 75:166-171 (93).		
gene	Location/Qualifiers		
CDS	1..1236		
	/organism="Arabidopsis thaliana"		
	/strain="L. Columbia"		
	/sub_strain="wild type"		
	/db_xref="taxon:3702"		
	35..1096		
	/tissue_type="whole plant"		
	35..1096		
	/gene="prxcb"		
	35..1096		
	/gene="prxcb"		
	/EC_number="1.11.1.7"		
	/codon_start=1		
	/product="peroxidase"		
	/protein_id="CAA50677.1"		
	/db_xref="GI:405611"		
	/db_xref="SPTRMBL:Q42364"		
	/translation="MHFSSSTSTSWTWLTLLGLMLHSLASIAQLTPFYDRSCPWV		
	TNIARETVNLEKSDPRIIASILAHHDLCFVNCGDASILLDNTPSRTEKDRGNAN		
	SARFEPIDRKKAVERACPRTVCADMLTLAAOOSYTLAGGPEMRVPLGRDLDAIF		
	LELANANLPAPFTFLPOLKASFRNVGDPSDLVALSGHTFGKNOCOFILDRPYNFS		
	NTGCLPTLNITYIQTLRGLCPRLNGNSALVDPUPLPYFDNKTYNNLEKERGLIOS		
	DOELFPSPNATDTPLVRAIRADGQTÖTFNFVEVMNRMGNTPTTGÖGOIRLMCRVV		
	NSNLHDVDVIDIVDFVSSM"		
	1144..1149		
polya_signal			

```

BASE COUNT      376 a      282 c      229 g      349 t
ORIGIN
/notes="basic subunit"
Query Match      53.2%; Score 493.6; DB 7; Length 1236;
Best Local Similarity 70.9%; Pred. No. 5.7e-139;
Matches 655; Conservative 0; Mismatches 269; Indels 0; Gaps 0;

4 caattaacccctacattctcagcaatagctgtcccaacgtgtcccaacatcgttcgagac 63
Db 125 CAACCTACACCCCTTACCTTTACGATAGTGTCTATGCTTATCTACATCACTGATGAGAA 134
QY 64 acaatcgtcaacagctcagatccgaatcccaagatcgtctcaatatactacgtcagac 123
Db 185 ACCATTGTAATAGATTAAAGGTGAGACCCCTGATGCGCTGCGAGCATCCCTGCTTACAC 244
QY 124 ttccatgactgctcgttgaaatggtgtgcagcgtatgacatatactatgacaacaccagt 183
Db 245 TTCACAGACGCTTTGTTAATGGTTTGACGATCCATCTTGTAGACACACGACATCA 304
QY 184 ttccgcacgtgaagaagatgcatcgcggaacgtcaacagcgccggggtcttcagtgatc 243
Db 305 TTTCGACACAGAGAAAGATCGGTTTGCAACGCAAAATTCGCTCGGGGATTTCCAGTGATT 364
QY 244 gatcgcatgaagagctgcgttgatgacagatgcccacgaacagtcagttgtgcagacctg 303
Db 365 GATAGATGTAAGACCTGCGGTGAGAGAGGCGATGCCCAAGACCGTTTCATGCGCAGATATG 424
QY 304 ctgacatagctgtgcgaacagagcgtgactctgcagcgagcgacccgttcgtgaagtgccg 363
Db 425 CTCACATTCGACCTGCAACATCTGTCTCTTGGAGAGGTCCTTGTGGAGGTTTCT 484
QY 364 ctggtgcagcgttactccctacagcatctcctagatctgccaacgccaactgctgct 423
Db 485 TTGGGAAGAGAGACAGATTTCACAGATTCCTGGAACCTGCTATGCAAAATCTTCCAGCT 544
QY 424 ccaatctcaaccctgcgccaacgtgaagatagctcttgaagaacgttggtctgaatctgctg 483
Db 545 CCATCTTTACACTTCACACTCCACAACTTAAGCCAGCTTCAGAAATGTTGTCGATCGTCT 604
QY 484 atgacactggtgtgctctgtgcggaagacaacatcttggaagaacagtgatgcatc 543
Db 605 TCTGATCTGCTTCTCTCTCTCGGTGTCACACATTTGGTAAATCATATGTCAGTTATT 664
QY 544 atgcatagctctacaaattccagcaacactggtgttaaccgtgaccccgctgaacactagc 603
Db 665 CTTCGACAGATTCTACAAATTTACAGAACACAGGTTTACCCGACCTTACACTCAACTACT 724
QY 604 tatccagaacactgagagcgtgtgcccactgaatgcaactcagtgactagtgagac 663
Db 725 TACCTCAAAACTCTGCTGATGATGCCCCCTTAATGGCAATGAGATGCTTGGTAGAT 784
QY 664 ttgtatctgcggagccccaacatctcgtataaagaatgactatgtgaatctagaagagcag 723
Db 785 TTTGATCTACGTAAGGCTTACGCTTTCGACAAACAAATACTACTGATGATCAAGAGCGA 844
QY 724 aaagggcctatacagagtgatcaagaactgtttaagcagtcacagccacactgacacatc 783
Db 845 AAAGGCTTATTCAGAGCGACCAAGAGTTGTTCTTAGGCCCAATGCGCATGACACAAATC 904
QY 784 ccaatgttgagaagatttgcttaactcactcaaaccttcttlaacgctctgttgaagcc 843
Db 905 CCGTTGGTAGAGCATATGCTGATGGCACAACAATTCCTTCAATGCATTTTGGAGGCA 964
QY 844 atggaccgtatgggttaacattaccctctgacgggtgaaccaagccagatgtctggaac 903
Db 965 ATGAATAATGATGGGAAGATTTACACCACTACAGGAATCMAAGACAAATACAGATTGAAC 1024
QY 904 tgcagagtggtcaacagcaactct 927
Db 1025 TGTAGAGTTGTGTAACCTCAACTCT 1048

```

```

RESULT 10
HRAHRPC
LOCUS      HRAHRPC      1229 bp      DNA
DEFINITION A.rusticana peroxidase isoenzyme C (HPR C) gene, complete cds.
ACCESSION M60729
VERSION    M60729.1 GI:168244
KEYWORDS   horseradish peroxidase.
SOURCE     A.rusticana
ORGANISM   A.rusticana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
            Armoracia.
            1 (bases 1 to 1229)
REFERENCE  1
AUTHORS    Fujiyama,K., Takemura,H., Shibayama,S., Kobayashi,K., Choi,J.-K.,
            Shimmyo,A., Takano,M., Yamada,Y. and Okada,H.
TITLE      Structure of the horseradish peroxidase isozyme c genes
JOURNAL    Eur. J. Biochem. 173, 681-687 (1988)
MEDLINE    88225087
FEATURES
            location/Qualifiers
            source      1..1229
                        /organism="Armoracia rusticana"
                        /db_xref="taxon:3704"
                        1..169
                        /gene="HRPC3"
                        /number=1
                        join(2..169,170..361,362..530,531..1000)
                        /gene="HRPC3"
                        join(2..169,170..361,362..530,531..1000)
                        /gene="HRPC3"
                        /codon_start=1
                        /protein_id="AAA33379.1"
                        /db_xref="GI:168245"
                        /translation="MHASFSAQLPTFFDSCRNVSNIYVDIIINELRSPSTIAS
                        ILRLHFHDFVNGQDSIILDNTTSFRLEKDAFGNANSARGPVVDRIKAAVERACPR
                        TVSCADVLITIAQDSVNLGDPMSRVPPLERDSRQAFDLANANLPAPEFTLPELKA
                        FANVGINRPSDVLVSLSGHTFEGNOCREFIMDLVNFSTGLPDLPTLNTYTLQTLKQC
                        PKNQGSVLYDEDLRTPVFDKRYVNLKEQGLIOSDELFSNPATPTIPLVSYA
                        DETQFFNFNFVFAAMRMGMNITPLUTGTGSRINCRVYNSLIHLIVEVDVSSM"
                        170..361
                        /gene="HRPC3"
                        /number=2
                        362..530
                        /gene="HRPC3"
                        /number=3
                        531..1000
                        /gene="HRPC3"
                        /number=4
BASE COUNT      337 a      275 c      256 g      361 t
ORIGIN
Query Match      52.9%; Score 490.4; DB 7; Length 1229;
Best Local Similarity 70.7%; Pred. No. 5.4e-138;
Matches 653; Conservative 0; Mismatches 271; Indels 0; Gaps 0;

4 caattaacccctacattctcagcaatagctgtcccaacgtgtcccaacatcgttcgagac 63
Db 29 CAACCTACACCCCTTACCTTTACGATAGTGTCTATGCTTATCTACATCACTGATGAGAA 88
QY 64 acaatcgtcaacagctcagatccgaatcccaagatcgtctcaatatactacgtcagac 123
Db 89 ATCATATTATCAATAGATTAGAGATGAGACCGCTAGATATGCGCGGACGATTCCTGCTTAC 148
QY 124 ttccatgactgctcgttgaaatggtgtgcagcgtatgacatatactatgacaacaccagt 183
Db 149 TTCACAGACGCTTTGTTAATGGTTTGACGATCCATCTTGTAGACACACGACATCA 208
QY 184 ttccgcacgtgaagaagatgcatcgcggaacgtcaacagcgccggggtcttcagtgatc 243
Db 209 TTTCGACACAGAGAAAGATCGGTTGGAACGCAACTCGCTGAGGATTTCTGTGTGTT 268

```

OY	244	gattcgatgaagagctgcgtttgagtcagcatgtgcccacgaaagctcaagttgtgagagactc	303
Db	269	gACAGATTAAGGCCCGCCGGTGGAGAGGCACTGCCAAGAATCTTTTCATCCGACGAGATGTG	328
OY	304	ctgacatagctgcgcgaacagagcgtgacatctgtcagagcgagacgcttccttgagagatgcg	363
Db	329	CTTACCATTTGCAAGCTCAACAACTCTGTTAATTGTGGCAGAGAGCTCTTCTTGAGGGTTCTCT	388
OY	364	ctcggltcgagcgtgtaacctccctacagagcatcttcagatcttggtccaaagccaaactgtgcgt	423
Db	389	TTGGGAGAGAGAGACAGCCGACAGACGATTTTGTAGATCTCGCTAATCGGAATCTTCCAGCT	448
OY	424	ccattcttaacctgtcccacagctgaagagtagctttagaagacgttggtctgcgaatcgtcg	483
Db	449	CCATCTTTTACACTTCCAGACACTTAAGCGTCTTTTGTCAATTTTGCCCTCAACCGCTCT	508
OY	484	agtgcacctgtgctctgtctgcgagagacacatcttggaagaagccagtgtagttatc	543
Db	509	TCTGATCTTCGTTCCTCTCTCTGTGGTGCACACATTTTGTGTAATAATCAATCTCGATTTATT	568
OY	544	atgagtagcgtctacaattccacacacttggtlttaactgaccccaagctgaaactagc	603
Db	569	ATGACAGATTTATACAACTTCAGCAACACCGGACTACCCGACCTACCCCTCAACACTACT	628
OY	604	tattccagacacactgagagagctgtgtcccactgaaatggaacactcagtcagtagtgac	663
Db	629	TACCTTCAAACTCTTCTGTCACACATGTCCCGTAAAGTGAACMAAGCGCTTGTGGAT	688
OY	664	tttgatctgcgagaccccaacatctctgataaagaatgactatgtgaaatctagaagagcag	723
Db	689	TTTCGATCTTCGTACGCCCAACAGTTTTCGATTAACAAATACTATGTGAATCTTAAAGACAA	748
OY	724	aaaggccctgatalcaagatgatalcaagaacactggtttagagatgcagagccacatgacatc	783
Db	749	AAAGCTCATCATCCAGAGTGACCAAGAGTGTGTTCTGTAAGCCCAATGGCACTGACACATC	808
OY	784	ccaactgtgagagaagtttctctaacttactactaaaccttctaagcgttcgttgaaagcc	843
Db	809	CCCTTGCTGAGATCATATATCTCTGATGGCACACAAACATCTTTCATGCTTGTGGAGGCC	868
OY	844	atggaccgtaabggtaaacatccctctcgaagcggtacccaaagccaaagatctgtcgaac	903
Db	869	ATGAATAGGATGGGAACATTTACACCTCTTACAGGAATCAAGAGAGAAATCAGATGTGAAC	928
OY	904	tgcagaatgtcacaagacaactctc	927
Db	929	TGTAGGGTGTGATCACTCAACTCT	952
RESULT 11			
ATPEROX5			
LOCUS	1195 bp	MRNA	PLN 07-OCT-1996
DEFINITION	A. thaliana mRNA for peroxidase ATP16a, EST clone 91D2217.		
ACCESSION	X98777		
VERSION	X98777.1 GI:1429220		
KEYWORDS	peroxidase.		
SOURCE	thale cress.		
ORGANISM	Arabidopsis thaliana		
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; Caprales; Brassicaceae; Arabidopsis.			
1 (bases 1 to 1195)			
Weininger, K.G., Jespersen, H.M., Kjaersgaard, I.V.H., Justesen, A.F., Oestergaard, L., Abelskov, A.K., Jensen, R.B., Hansen, L.N. and Rasmussen, S.K.			
From expressed sequence tags to structure, function, evolution and expression of 28 ER-targeted Arabidopsis peroxidases unpublished			
2 (bases 1 to 1195)			
Justesen, A.F.			
Direct Submission			
Submitted (25-JUN-1996) A.F. Justesen, Institute of Molecular Biology, Department of Protein Chemistry, Oester Farimagsgade 2A,			

FEATURES									
DK-1353 Copenhagen K, DENMARK									
Location/Qualifiers									
Source									
1. .1195									
/organism="Arabidopsis thaliana"									
/cultivar="Columbia"									
/db_xref="taxon:3702"									
/tissue_type="stems, leaves, flowers, roots and siliques"									
/clone_1lb="Lambda PRL2 (Dr. T. Newman et al)"									
/clone="EST clone 91D22n7"									
32. .1118									
32. .1090									
/EC_number="1.11.1.7"									
/codon_start=1									
/product="peroxidase App16a"									
/protein_id="CA67313.1"									
/db_xref="GI:1429221"									
/db_xref="SPRMBL:O43732"									
/translation="MNFSSYSLSIWTTLMTLGLCLLHLSSAQLPFPFDNTPSYF TIVVDLNEKSDPRILAASTLRIHFDCEVSGCDASILIDNTTSPFEXDAPNAMS ARGPVIVDRMKAAYETACPRVSCADLILVAQDVNADLWAGPSMRVPLGHRDLOAF ALANPLPAPEFTLPOLKASFORNGVDRPDLVLSGHTFGKQCFIDRLYRNSN TGLDPLNTLYLQTLRGOCPRNGNORVLDVDFLDRPTVFDNNYVVKELKELIOTD OELSSPNADTDLPIVREYADGQKFPNAVEANNRGNITPLTGQGIKONCRVYN SNSLIHVEYLVDPVSSM"									
mat_peptide 119.1042									
BASE COUNT 354 a 278 c 223 g 340 t									
ORIGIN									
Query Match 51.9%; Score 481.4; DB 7; Length 1195;									
Best Local Similarity 70.1%; Pred. No. 2.9e-135;									
Matches 647; Conservative 0; Mismatches 276; Indels 0; Gaps 0;									
QY	4	cagtaacccccacac	ttctagcaaatagctg	gtcccaacgtg	ccaacatggtc	ggac	63.		
DB	119	CAGCTAACCCCACT	TTCTATGACATACAT	GTCCAACGCTT	TTACCATCGTAC	GGGAT	178		
QY	64	acaatcgtcaacag	ctcagaatccgatcc	ccagatcgc	gtctcaata	ctatcgtc	gac	123	
DB	179	ACCATAGTAAAG	CAGACTAAGATCGG	ATCTCGTATCG	TGCAGACATCTT	CGCTTTCAC	238		
QY	124	ttcgaatcgtc	ctctgtaatggtgtg	cgacgtatgaca	tatatctactg	tcgacaacacacac	gac	183	
DB	239	TTCCACGACTGT	TTCGTAATGGCTGT	ATCATCGATCC	TGTTAGAACACAA	CATCC	298		
QY	184	ttccgacttgaa	agatgcatcgtcg	ggaagctaa	cagcgccca	gggtcttcc	gtagtc	243	
DB	299	TTTTCAACAGAA	AGATGCGGCTCCA	ACCCAACTCG	CTCGAGATTC	CGAGTAT	358		
QY	244	gatcgatgaag	gctgcgcgtttgatg	tcagcatcgtc	ccagcaaac	gaatgagtcg	gac	303	
DB	359	GATTAATGAAT	GAAGCAGCAGTAG	AGACACATGTCC	AAAGACTGTT	TCATGCGCAGAT	ATA	418	
QY	304	ctgactatagct	gtcgcaacagagc	gtgactcttc	gagcgga	ccgtctctg	agagtc	363	
DB	419	CTTACCATTCG	CAGCTCAACAAGCT	TAATAATTGG	CAGAGGTC	TTCTTTGGAGGGT	TCT	478	
QY	364	ctcgctgacgt	gtactccctctaca	ggcatcttct	gagtcgg	ccaacgcca	acttgct	423	
DB	479	TTGGGGAAAG	AGAGAGCTTTACA	AACTATTTTGT	CTTCGCTAA	ATACAAATCTT	CCCCCT	538	
QY	424	ccatcttcaac	ctcccccaagct	gtgaagatag	cttga	aaacgttg	gtctga	483	
DB	539	CCATTCCTTCA	CTCTCCCACTTA	AGCCAGT	TTTCAAAATG	TTGGAGCTTG	ACCGCTCC	598	
QY	484	agtgacctgt	gtcctgtccg	gaggaacac	atlttga	aaagaa	ccagtc	543	
DB	599	TCTGATCTTC	GTGTGCACTCTG	TGTGTGTC	ACACATTTGG	TAAAAACCA	CTGCCAATTTAT	658	
QY	544	atgagtaagc	ctctacaatltt	cagaacac	ctggttact	ctgccccca	gcgtga	603	
DB	659	ATGGATACAT	ATACACCTTTAG	CAACACTGG	TTTATCCG	CACTTACTT	TAATATCA	718	





Db 835 AAGGACCTTATTCAGACGATCAAGAGTTGTTTTCAGCCCTTAATGCACTGACACATC 894  
QY 784 ccaatgtgtagaagtttcttaactactcaaaccttctaagcgttcgtggaagcc 843  
Db 895 CCTTGTGTGAGACATATACCGTGTGACTCAAAAATTTCTCAATGCTTTTGTGGAAGCA 954  
QY 844 atggaccgtatggtaacattaccctctgaaggttaaccgaagccagatctgtaaac 903  
Db 955 ATGAATAGATGGGAACATTACACTCTCCTGGAATCAAGACAAATACAGAGAAT 1014  
QY 904 tgcagagtggtcaacagcaactc 926  
Db 1015 TGTAGGTGTGTTAACTCAACTC 1037

RESULT 13  
PPXP2PER  
LOCUS PTPXP2PER 1202 bp mRNA PLN 12-MAY-1997  
DEFINITION P. trichocarpa mRNA for anionic peroxidase Pxp22.  
ACCESSION X97349  
VERSION X97349.1 GI:1279649  
KEYWORDS peroxidase.  
SOURCE Populus balsamifera subsp. trichocarpa.  
ORGANISM Populus balsamifera subsp. trichocarpa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
Rosidae; Violales; Salicaceae; Populus.  
REFERENCE 1 (bases 1 to 1202)  
AUTHORS Christensen, J.H., Bauw, G., Boerjan, W. and Van Montagu, M.  
TITLE Purification and cloning of poplar xylem peroxidases  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1202)  
AUTHORS Christensen, J.H.  
TITLE Direct Submission  
JOURNAL Submitted (19-Apr-1996) J.H. Christensen, University of Gent,  
Laboratory of Genetics, Ledeganckstraat 35, Gent, B9000, Belgium  
FEATURES  
Location/Qualifiers  
Source 1..1202  
/organism="Populus balsamifera subsp. trichocarpa"  
/sub\_species="trichobol"  
/db\_xref="taxon:3694"  
/clone="Pxp22"  
/issue\_type="xylem"  
10..1041  
/EC\_number="1.11.1.7"  
/function="signal for ER"  
/codon\_start=-1  
/product="peroxidase"  
/protein\_id="CAA66035.1"  
/db\_xref="GI:1279650"  
/db\_xref="SPTREMBL:Q43100"  
/translation="MHISKAIVEAFEFVVLRLNGTYLAGCOLPTPEYDQCPNVSSIIRD  
VTEFLVSDPRIGASILRLHFDPCVNGCGSLLDNDPTVISEKAGGNNSARFE  
VVDTKALLESACPAIVSCADITLIAEESVVLGAPNTVPLGRDSTYASRDANA  
FLPAPFETLDQRESFTNLSNNSDVLVSGAHFGRKSGTFPDRLYDFNSTGAPD  
PSLDITFLAALQELCPGSGNSVITDLDISTPDADSDVYSNLQNGRGLQPDQELS  
TPGADVIALVVAFSANQTAPESEFVESHIRMGNTLSPLTGTGEBIRLNCVYVANNLAG  
PDSMLVSSI"

BASE COUNT 348 a 240 c 269 g 345 t  
ORIGIN

Query Match 32.8%; Score 304.4; DB 7; Length 1202;  
Best Local Similarity 58.1%; Pred. No. 1.8e-81;  
Matches 536; Conservative 0; Mismatches 386; Indels 0; Gaps 0;

QY 4 caatlaaccctatctcaacaatagctgcccaagctgccaacatcgcttcgcgac 63  
Db 82 CACCTTACTCTCAACATTTTATGACACACATGTCCAAATGTGACACCATATTCCTGGAT 141  
QY 64 acaatcgtaacagctcagatccgatacccaagatcgctctcaatatactgctgcac 123  
Db 142 GTCATTACGAGACATTGTAATCCGATTCCTGGATTGAGCCAGCCTCATCAGGCTCCAC 201

QY 124 ttccatgactgcttcgtgaatggctgcagcctagcalatlaactgacaaaccaccagt 183  
Db 202 TTCCATGACAGCTTTTGTATATGGCTGTGATGTTCCCTTTTGTGGACATATGATACT 261  
QY 184 ttccagactgaagaagatgcatctcggaacgctaacagcgccaggggtcttccagtatc 243  
Db 262 ATAGTAGGCAAAAGAAAGAACGCCGAGGAGAACACACTCTCAGAGAGTTTGAAGTTGTT 321  
QY 244 gatcgatgaagctgcgcgttgatgtaagcattgcccaagcaacagtcagttgtgagacctg 303  
Db 322 GATACATGAAAGGCTTTGTTGGAGAGTGCCTGCTCACTGTTCTCTGCTGTGATATA 381  
QY 304 ctgactaagctgcgcaacagcagcgtgactctgcgagcgaccgctccctgagatggccg 363  
Db 382 CTCACATTTTCACCTGACATGACATGCTGTTTGTGGACAGAGTCCAAATTTGGACAGTTCCA 441  
QY 364 ctgcgtcagctgactccctcactcagagcattcctagatctggccaacgccaacttgctgct 423  
Db 442 TTAGGAAGAAAGATAGACACACAGCAGCAGATGACCAATATGTTTCCAGGCC 501  
QY 424 ccaattcaccctcgcccaagctgaagatagctttagaagcgtgggtctgtaactgcctcg 483  
Db 502 CCTTTTTCACCTTGATCAACTCAGAGAGGCTTCACTAATGTCAAGCCTTAATAATAAT 561  
QY 484 agtgaacttggtgctctgctgcggaagacacacattggaaagacagctgtaagttcactc 543  
Db 562 AGTGATTTGTACTCTATCTGTGTGCTCACACATTTGGGGGCAAAATGTTCTACATTC 621  
QY 544 atgatatgctctcacaatttcaagaacactgggttcaactgaaccacagctgaacactagc 603  
Db 622 GACTTCGATTTGATATATTTCAACAGACCGGTCCTGACCATTCGTTGGACACACT 681  
QY 604 tatctccagacactggaagctgctgcccactgaatggcaacctcagctagctagtgagc 663  
Db 682 CTCTTAGACACTTTCAGGATTAATGTCGCCGAAGGTGAATGGAGTGTGATACAGAT 741  
QY 664 ttgatctcgagaccccaacactctcgtatcaacaagtaactgtgaactgaaggagagcag 723  
Db 742 CTTGATCTTCACACACTTGATGCTTTTGACAGTGTACTACTTCACATTCGACAGGTAAC 801  
QY 724 aaagcctgatacagaatgatacagaactgtttagcagtlcccaagcccaactgaaccatc 783  
Db 802 CGAGGCCCTGTTCAACTGATCAAGAACTGTTTCAACTCGGGGGGAGATGATGTGAT 861  
QY 784 ccaatgtgtagaagtttcttaactactcaaaccttctaagcgttcgtggaagcc 843  
Db 862 GCACCTGTTAACGCTTTCAGTGTCAATCAAAACAGCTTTCTTGAAAGCTTTGTGAGATCC 921  
QY 844 atggaccgtatgggtatacatcattaccctctgaaggggtlaaccaagccagatctgctgaac 903  
Db 922 ATGATAGAAATGGGAATTCACAGCCCTCTGACAGAACTGAAGGAGATCAGATTCAC 981  
QY 904 tgcagagtggtcaacagcaactc 925  
Db 982 TGCAGCGTACTCAATGCAAACT 1003

RESULT 14  
POPP01  
LOCUS POPP01 1143 bp mRNA PLN 06-FEB-1999  
DEFINITION Populus nigra mRNA for peroxidase, complete cds.  
ACCESSION D83224  
VERSION D83224.1 GI:1199775  
KEYWORDS peroxidase.  
SOURCE Populus nigra (variety: Italica Koehne) cDNA to mRNA.  
ORGANISM Populus nigra  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
Rosidae; Violales; Salicaceae; Populus.  
REFERENCE 1 (bases 1 to 1143)  
AUTHORS Sakuma, Y.  
TITLE Direct Submission

REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES
JOURNAL	Submitted (26-JAN-1996) to the DBJ/EMBL/GenBank databases. Yoh Sakuma, Institute of Wood Technology, Arita Prefectural College; 11-1 kaiezaka, Noshiro, Akita 016, Japan (E-mail:yakuma@agr.hokudai.ac.jp, Tel:0185-52-6984, Fax:0185-52-6976)			
2 (bases 1 to 1143)	Sakuma,Y., Azuma,T., Kato,Y., Kojima,Y. and Miura,K.	Poplar peroxidase gene responding to cytokinin treatment	Unpublished (1996)	
Location/Qualifiers				
1..1143	/organism="Populus nigra"			
	/variety="italica Koehne"			
	/db_xref="taxon:3691"			
1..>1143				
9..1040	/codon_start=1			
	/product="peroxidase"			
	/protein_id="BA11852.1"			
	/db_xref="GI:1199776"			
	/translation="MHISKAIVAAFFVYLIGTLACQLTPRYDQCPNVSSIRIRVITTELVSDPRIGSLRLRHPDCLVNCDDSLLDNDITVSEKAGNNNSARGFELVDDMKALIESACPAIVSCADILITAAESVVLGAGPMWPLRLRSTSRPAANFLDPKFLFDLDRSEFTVSNNTMDVLAISGAHFGAKSTPDRFLFENSGAPRQSLMTLLADLOELCPGSGNSVITDILDTTPDPAFDSVYSYINLGNGSLIOTDELFTPGADVDVIAFNATFESFESANOTAFEFSEFESMIMGNLSPLTGEIRLRCRVNANLAGPDSMLVSSI"			
BASE COUNT	319 a	236 c	261 g	327 t
ORIGIN				
polya_signal				
Query Match	32.6%;	Score 301.8;	DB 7;	Length 1143;
Best Local Similarity	58.0%;	Pred. No. 1.1e-80;		
Matches 534;	Conservative 0;	Mismatches 387;	Indels 0;	Gaps 0;
4	cagtaaaccccaactctcagacaaatagctgtcccaacggtgccaaatcgttcggac	63		
81	cc	140		
64	acaatcgtaacagagctcagatccgatacccaagatcgctgtcctaataatcgtcac	123		
141	gtcatttaggagacatttgatccatctctcgatttgagaccagctctatcagctccac	200		
124	ttccatgactgcttcgtgaatgtgttcgacgctagacataatccttgacaacacacag	183		
201	ttccatgactgcttcgtgaatgtgttcgacgctagacataatccttgacaacacacag	260		
184	ttccatgactgcttcgtgaatgtgttcgacgctagacataatccttgacaacacacag	243		
261	atagtagacggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga	320		
244	gacgcatgaagagcgcgcgttgatgcagacatgcccaagacagctaatgtgtcagac	303		
321	gattagaaatgaagagcgtttgttgagagtgctctgtctgactgtttccctgcatata	380		
304	ctgactatagctctgcgaacagagctgtgactctctgcaggcgcgacgctccctgagag	363		
381	ctcactatgtgcagctggaagaaatctgttgcctggcagagagctcgaatgttgaca	440		
364	ctcggctgacgtgactccctacaggaactctagatctgtgccaacgcaactgtcgtc	423		
441	tttagcaaacagaggaatgacaaacagacccagatgacgaacaaatgctttcccttc	500		
424	ccattctcacccctgcgccagctggaagatgacttagaagcgtgtgtgtcgaatcgc	483		
501	cttttttttcccttgatcaactcagaagagccttacctaatgtacgcttataatata	560		
484	agtagcctgt	543		
561	actgattttgtgtagctctatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	620		
544	atggaatgcttcaacttcaagcaacatgtgttacttcgaccccaacgctgaacatc	603		

	DB	621	GACCTCCGATTGTGGATTTTCAACAGCACCGTGCTCCTGCACAATCATCTGAACACACT	680
OY	604	tatctccgaacactgagaggcttgtgccactgaattggcaacctcgcagttagtgag	663	
Dd	681	CYTTCAGCAGATCTTCAGAGAATTAATGATGCCCCCAGAGTAGAAATGGAGAGGTATAAAGAT	740	
OY	664	tttgttcgtcgagaccaccaacalcitlgataacaagaictctyfgaatcatvtagagagcacg	723	
Dd	741	CTTTGATCTTCACAAACCCTGANOCCTTTTACAGTAACTPCTTACTCCAAATCTCCAAGGTMAAC	800	
OY	724	aaaagccgatatacaaagatgatalcaagaactgtltagcaatlccaagacyccacatgacaacatc	783	
Dd	801	CNAGGCGTCTTCACACTGATCATAGAATCTGTTTTCACTCTCTGGGCGCAGATGATGTCTATT	860	
OY	784	cacactgytgaagaasfittgtctaactctactcataaacctcttaaacgcctctgttgaagcc	843	
Dd	861	GCACCTGTGTAACGCTTTTCAGTTCATGCTAATCAAACAGCTTCTTTGAAAAGCTTTGGCGAGTCC	920	
OY	844	atygaccgttatvggtatacaltaeccctctgcagggtlaccagaagccagatlcgtctgac	903	
Dd	921	ATGATTAACAATATGGGAAAACTCAGCCCCCTGTGACAGGAACACTGAAGAGAGAGATCAGATTGMAC	980	
OY	904	tgcagagatggtccaacagcac	924	
Dd	981	TGCAGGGTAGTCAATGCAAAC	1001	
RESULT#	15			
PTRXPAPER	LOCUS	PTPXPPER	1371 bp mRNA	PLN
DEFINITION		F.Trichocarpa mRNA for anionic peroxidase Pxpl.		12-MAY-1997
ACCESSION	X97351			
VERSION	X97351.1	GI:1279653		
KEYWORDS		peroxidase.		
SOURCE		Populus balsamifera subsp. trichocarpa.		
ORGANISM		Populus balsamifera subsp. trichocarpa. Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; Violales; Salicaceae; Populus. 1 (bases 1 to 1371)		
REFERENCE	AUTHORS	Christensen,J.H., Bauw,G., Boerjan,W. and Van Montagu,M.		
JOURNAL	TITLE	Purification and cloning of poplar xylem peroxidases Unpublished		
REFERENCE	AUTHORS	Christensen,J.H.		
TITLE		2 (bases 1 to 1371) Direct Submission Submitted (19-APR-1996) J.H. Christensen, University of Gent, Laboratory of Genetics, Ledeganckstraat 35, Gent, B9000, Belgium		
JOURNAL		Location/Ounaliers		
FEATURES	source	1..1371		
	/organism="Populus balsamifera subsp. trichocarpa"			
	/sub_species="Trichobel"			
	/db_xref="taxon:3694"			
	/clone="Pxpl"			
	/rname_type="xylem"			
	54..1091			
	/EC_number="1.11.1.7"			
	/function="signal for ER"			
	/codon_start=1			
	/product="peroxidase"			
	/protein_id="CA66037.1"			
	/db_xref="GI:1279654"			
	/db_xref="SPRMBL_043102"			
	/translation="MTDKALRLPVASLFPIYWFGSGSLPYAYADLTPEFDICPNVST IRKVLQALQTDRIGASLRHLRFHDCFPVGSDGSILTNDNTITSEKEEAANNNSA RGSDVDNNKAAMAVENACPGISCADILAIABESVRLAGPSVTVPJGRDLSIARSS GNASSIPAPSESIALSKPAVAAGLANSSDLVALSCAHFPGRAOCLNFISRLXFFSGS GNDDPTLNTRYLAALOQLPOGGNGRSVLTNLDTTPPDTEPNGAFSNLOTFNEGLIOSDO ELNSTGDRTIAIVNNSSNMOTAFPFEFSFYVMIRMCMNISLPLOTDEIRLNCRIVANS TGSALLLVSSI"			
CDS				
BASE COUNT	387 a	280 c	288 g	416 t
ORIGIN				





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2000, 06:40:28 ; Search time 101.52 seconds

(without alignments)  
2284.555 Million cell updates/sec

Title: US-09-246-451-16

Perfect score: 927

Sequence: 1 atgcagtaaccctacatc.....gagtgtaacagcaactct 927

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues  
Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_36.\*  
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	925.4	99.8	955	1 N90760	Synthetic horserad
2	922.4	99.5	1055	1 N90762	Synthetic horserad
3	843.8	91.0	1072	1 Q37614	Encodes HRP/IGF-II
4	563	60.7	927	1 N93078	Sequence encoding
5	525.6	56.7	924	1 N81508	Horseshadish peroxi
6	485	52.3	501	1 N90761	5' end half of syn
7	446.4	48.2	474	1 N94368	3' end half of syn
8	239.8	25.9	1244	1 V30460	Soybean seed coat
9	239.8	25.9	1244	1 V59077	Soybean seed coat
10	230.4	24.9	1355	1 V63720	Maize per5 root pr
11	163.4	17.6	1270	1 087101	The Tpa gene. Tran
12	163.4	17.6	1270	1 T63685	Full-length Tpa-1
13	163.4	17.6	1270	1 X03970	Rice green leaf pe
14	161	17.4	1348	1 Q39021	Full length anionl
15	159.4	17.2	388	1 Q37615	Encodes HRP/IGF-II
16	148.6	16.0	1131	1 V81611	Basic cucumber per
17	148.6	16.0	1131	1 V62837	Cucumber peroxidase
18	147.4	15.9	1131	1 006207	Cucumber peroxidase
19	146	15.7	1181	1 V81616	Cucumber peroxidase
20	146	15.7	1181	1 V62867	Cucumber peroxidase
21	129	13.9	1064	1 087093	Tpa-1 phage insert
22	129	13.9	1064	1 T63677	Peroxidase gene. T
23	129	13.9	1064	1 X03962	Antionic tobacco pe
24	119.8	12.9	1326	1 T66685	Soybean peroxidase
25	119.8	12.9	1326	1 V69274	Soybean peroxidase
26	119.8	12.9	1326	1 V81421	Soybean peroxidase
27	118	12.7	3359	1 V59078	Soybean seed coat
28	118	12.7	4700	1 V30461	Soybean seed coat
29	114.4	12.3	1144	1 T95782	Scylsanthines humil
30	107	11.5	1167	1 T66687	Soybean peroxidase
31	107	11.5	1167	1 V69276	Soybean peroxidase
32	107	11.5	1167	1 V81423	Soybean peroxidase
33	105.4	11.4	1315	1 T66684	Soybean peroxidase

34	105.4	11.4	1315	1 V69273	Soybean peroxidase
35	105.4	11.4	1315	1 V81420	Soybean peroxidase
36	98.2	10.6	1191	1 T66686	Soybean peroxidase
37	98.2	10.6	1191	1 V69275	Soybean peroxidase
38	98.2	10.6	1191	1 V81422	Soybean peroxidase
39	88	9.5	6550	1 V63717	Maize per5 root pr
40	79.8	8.6	686	1 V81617	Cucumber peroxidase
41	79.8	8.6	686	1 V62868	Cucumber peroxidase
42	75.2	8.1	278	1 V23835	Plant POX enzyme D
43	71.4	7.7	687	1 V23906	Plant POX enzyme D
44	60	6.5	688	1 V23907	Plant POX enzyme D
45	59.6	6.4	1410	1 V23905	Plant POX enzyme D

## ALIGNMENTS

RESULT	1	
ID	N90760	
AC	N90760 standard: DNA: 955 BP.	
DT	18-JUL-1990 (first entry)	
DE	Synthetic horseradish peroxidase gene (HRP)	
KW	Horseradish peroxidase; restriction sites; DNA construct;	
KW	fusion protein; enzyme assays; immunoassays; synthetic gene.	
OS	Horseradish.	
FH	Key	Location/Qualifiers
FT	cds	11..937
FT		/*tag-a
FT		/note-bases 14-937 are claimed in claim 2"
PN	MO8903424-A.	
PD	20-APR-1989.	
PF	07-OCT-1988: G00833.	
PR	08-OCT-1987: GB-023662.	
PA	(BRBR-) Brit Bio-Tech Ltd.	
PI	Edwards RM, Burke JF;	
DR	WPI: 89-130042/17.	
PS	P-PSDB: P95200.	
PT	Synthetic gene encoding horseradish peroxidase -	
PT	used for efficient prodn. of the enzyme or fusion prods. for	
PS	use in enzyme assays, esp. immunoassays	
PS	Disclosure: 45pp: English.	
CC	The synthetic gene has many restriction sites. Also claimed is a DNA	
CC	construct contg. bases 14-937. The construct can be a vector where the	
CC	HRP gene and a desired gene are co-expressed as a single fusion prod.,	
CC	a single polycistronic message, or 2 separate but linked transcriptional	
CC	units. The fusion proteins are for use in enzyme assays, esp. immuno-	
CC	assays. Genes carrying mutations that destroy HRP activity can be used	
CC	to follow the frequency of reversion or suppression of the particular	
CC	mutation introduced into the gene. The constructs can be used in	
CC	transgenic animals, esp. mammals. The synthetic gene is prepd. by	
CC	coupling successive nucleotides and/or ligating appropriate oligomers.	
SQ	Sequence 955 BP; 246 A; 269 C; 221 G; 219 T;	
Query Match	99.8%; Score 925.4; DB 1; Length 955;	
Best Local Similarity	99.9%; Pred. No. 5.6e-291;	
Matches 926; Conservative	0; Mismatches 1; Indels 0; Gaps 0;	
QY	1 atgcagttaacccctacattctacagacaatactgtcccaacggtgccacaatcgttcgc 60	
Db	11 ATGAGTTAAACCCCTACATTCTAGACACATATACCTCTCCCAAGTGTCCCAACATCGTCCG 70	
QY	61 gaacacaatcgtaacagagctcagatccgatccacagagatcgctgtctcaataatactgctg 120	
Db	71 GACACAAATCGTCAACAGACTCGAATCCGATCCACAGATCGCTTCATATATTAGCTGCG 130	
QY	121 caattccatgactgcttgtaagtggttcgacagctatgacatattacttgaaacaccacc 180	
Db	131 CACTTCGATGACTGCTTGTGTAATGCTTGCCGACGCTAGCATATTACTGACACACACACC 190	
QY	181 agtttcgactgaaagaatgacatgacatgggaacgctaaacagcgcgaagggtcttcagtg 240	

```
Db 191 AGTTTCGGCACTGAAAGATGCATTGCGGAGACGCTAAACGCCGAGGGCTTTCAGTg 250
Qy 241 atcgatcgatgaagcgtgcgcttgatgcagatgcgcccagaacagtcagttgtgcagac 300
Db 251 ATGCATGCGATGAAAGCGTGGCGTTGATCAGCATGCCCCAGCAACATGATGTCGAC 310
Qy 301 ctgctgaactaagctgcgcacaacagcgtgactctgcagcgcgaccgtccctggaagtg 360
Db 311 CTGCTGACTATAGCTGCGCAACAGAGCGTACTCTTCAGCGCGGACGCTCCTGGAGAGTG 370
Qy 361 ccgctgctggaagctgactccctcaagaagcattcctatactgagcgaacgcgaacttgc 420
Db 371 CCGCTCGGTGACGTGACTCCCTAACAGGCAATCTCTAGATCTGGCAACGCCCACTTCCCT 430
Qy 421 gctccattcttaaccctgcgcccacgctgaagatagctttaagaacgctgggtctgaatcgc 480
Db 431 GCTTCATTCTTCACCTGCGCCAGCTTAAGATAGCTTTGAAACGTCGGGTCTGAATCGC 490
Qy 481 tcgagtgacctgtgtctctgtccgagagacacacatttgaaagaaccagtgtagtgc 540
Db 491 TCGAGTGAACCTTGTGCTGTGTCGAGAGACACATTTGGAAGAACAGTGTAGGTTTC 550
Qy 541 atcatgataagcgtctcaaatcttaagaacacactgggttactgaccccaagctgaacact 600
Db 551 ATCATGGATAGGCTTAACTTTCAGCAACACTGGGTTACCTGACCCAGCGTGAACACT 610
Qy 601 acgatctcagaacacactgaagagcgtgtgccactgtaatggcaacctcagtgacactgtg 660
Db 611 ACGATCTCAGACACTGAGAGAGGCTTGTGCCACTGATGATGGCAACCTCAATGCGCATG 670
Qy 661 gaacttgatctgcggagcccaacacatctcgataacaagtaactatgtgaactagaagag 720
Db 671 GACTTTGATCTGGCGGACCCCAACCATCTTGATCAAGTAAGTATGTGAATCTGAGAGAG 730
Qy 721 caaagaagcctgtatagagagtgatcaagaactgtttagcagtcgaagcgcacatgaaac 780
Db 731 CAGAAAGCCTGATACAGAGTGTATCAAGAACTGTTTAGCAGTCCAAAGCCACTGACACC 790
Qy 781 atcccaatgctgaagaatttcttaactctactcaaacctctttaaagccttcgtgaa 840
Db 791 ATCCACTGCTGAGAAATTTTGGTAATCTACTCAAACTCTTTAAAGCCTTGTGTGAA 850
Qy 841 gccatggaacgtatgggtaacataccctctgaacgggtaacccaagcgcaagatcgtctg 900
Db 851 GCCATGGAACGCTATGGGTAACTATACCCCTCTGAGCGGTACCCAGAGCGCATCTGCTG 910
Qy 901 aactgcagagtgtaaacagcaactct 927
Db 911 AACTGCAGAGTGTCAACAGCAACTCT 937

RESULT 2
N90762
ID N90762 standard; DNA; 1055 BP.
AC N90762;
DE 18-JUL-1990 (first entry)
DE Synthetic horseradish peroxidase gene (HRP) modified for expression
DE in mammalian cells
KW Horseradish peroxidase; restriction sites; DNA construct;
KW fusion protein; enzyme assays; immunoassays; synthetic gene;
KW synthetic linker; PSD18.
OS Horseradish.
FH Key
FH misc_feature Location/Qualifiers
FT 1..73 /*tag= a
FT /*note="linker used to adapt synthetic gene"
FT misc_feature 993..1055 /*tag= b
FT /*note="as above"
FT cds 12..68 /*tag= c
FT /*product=N-terminal pre sequence
FT 993..1040
```

```
FT /*tag= d
FT /*note="C-terminal pro sequence"
FT mat_peptide 69..992 /*tag= e
FT /*note="claimed in Claim 2"

WO8903424-A.
PD 20-APR-1989.
PF 07-OCT-1988. G00833.
PR 08-OCT-1987. GB-023662.
PA (BRRI-) Brit Bio-Tech Ltd.
PI Edwards RM, Burke JF;
DR WPI: 89-130042/17.
DR P-PSDB; P95203.
PT Synthetic gene encoding horseradish peroxidase -
PT used for efficient produ. of the enzyme or fusion prods. for
PT use in enzyme assays, esp. immunoassays
PS Disclosure; ; 45bp; English.
CC The synthetic gene has many restriction sites. The synthetic gene is
CC prep'd. by coupling successive nucleotides and/or ligating
CC appropriate oligomers. The front end of the synthetic HRP gene was
CC modified by the replacement of the HindIII-HpaI fragment (see n90760)
CC with a synthetic linker carrying an NdeI site on the initiator ATG. The
CC synthetic HRP gene was cloned into the expression vector pBC517 on a
CC NdeI-BamHI fragment to give the plasmid PSD18. The plasmid was maintained
CC E.coli strain W3110 lacIq and high level of expression of active HRP
CC was obt'd. A construct containing the claimed sequence (See FT) is
CC claimed. The construct can be a vector where the HRP gene and a
CC desired gene are co-expressed as a single fusion prod.
CC a single polycistronic message, or 2 separate but linked transcriptional
CC units. The fusion proteins are for use in enzyme assays, esp. immuno-
CC assays. Genes carrying mutations that destroy HRP activity can be used
CC to follow the frequency of reversion or suppression of the particular
CC mutation introduced into the gene. The constructs can be used in
CC transgenic animals, esp. mammals.
SQ Sequence 1055 BP; 260 A; 294 C; 250 G; 251 T;

Query Match 99.5%; Score 922.4; DB 1; Length 1055;
Best Local Similarity 99.9%; Pred. No. 5.5e-290;
Matches 923; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 caglttaaccctaatctctacgaacaatagctgtcccaagctgtccaacatcgttcgac 63
Db 69 CAGTTAACCCCTTACATTCTACGACAATATAGCTGCCAACTGTCCAACTATGCTTGGCG 128
Qy 64 acaatgcgaagcgtcgaatccgataccgagatgcgtcttcaataatagcttgac 123
Db 129 ACAATGCTCAACGAGCTCAGATCCGATCCAGATCCGCTCTTCAATATAGCTTGCAC 188
Qy 124 ttccatgactgcttgtaatgtgttgacgctagcatatctactggaacaacacagct 183
Db 189 TTCCATGACTGCTTTCGGAATGTTGTGAGCGTATAGCATTTACTGGAACAACACCAC 248
Qy 184 ttccgaactgaaagatgcatctcgggaacgctaacagcgccaggggtcttcagtgatc 243
Db 249 TTCCCGCACTGAAAGATGATTCATTCGGGAACGCTTAACAGCCGAGGGGCTTCCAGTATC 308
Qy 244 gatcgaatgaagcgtccgttgatgcaagcctgcccagcaactgattgtgagagctg 303
Db 309 GATCGCATGAAGGCTGCGTGTGAGTCAGCATGCCACAGAACATGCTATTTGTGAGACTG 368
Qy 304 ctgactataagctgcgaacagagcgtgactctgcagcgcggaacgctccttgaaagtgcg 363
Db 369 CTGACTATAGCTGCGCAACGAGAGCGTACTCTTGAAGCGGACCGTCTGGAAGTCCG 428
Qy 364 ctcgatcgacgtgactccctacaggaatctcctagatctggccaagccaactgtcgtc 423
Db 429 CTCGCTGAGAGTACTCCCTTACAGGCAATCTCTAGATCTGGCAACGCAACTTGGCTG 488
Qy 424 ccaatctcaacccctgcccagcgtgaagatagctttagaagcgtggtcgaatcgtcg 483
Db 489 CCAATCTTACCCCTGCCCGCGTGAAGGATAGCTTTAAGAAACGTGGGTGATTCGCTCG 548
```

QY 484 atgagacctgtgctcgtcccgagagacacacattggaaagaacagtgtagtcaac 543  
 |||||||  
 Db 549 AGTCACCTTTGGCTCTGTCCGGAGACACACATTGTGAAAGAACACGTGTAGCTTCATC 608  
 QY 544 atgatatggtctacaatttcagacaacacgtggttacctaacccacagcgaacactacg 603  
 |||||||  
 Db 609 ATGGATAGGCTCTACATTTCAGACACACTGGGTTACTCTACCCACGCTGAACTACTACG 668  
 QY 604 tatctccagacactgaagagctgtgtcccaactgaatgcaacacctgacatgtacgtacg 663  
 |||||||  
 Db 669 TATCTCCAGACACTGAGAGAGCTTGTCGCCCACTGATATGCAACCTCATGTGACTGTGAC 728  
 QY 664 ttgatctgcgagaccaccaacatcttcgataaacaagtactatgtgaactagaggagacag 723  
 |||||||  
 Db 729 TTTCATCTGGGAGCCCAACCATCTTGTGATTAACAAGTACTATGTGATTAAGAGACAG 788  
 QY 724 aaaggacctgaagagtgatgttcaagaactgttagcagctcagacgacctgaacacatc 783  
 |||||||  
 Db 789 AAAGGCTGTATACGAGATGTATCAAGAACTGTTAAGCATCCAAACGCCACTGACACATC 848  
 QY 784 ccactgtcgagaagtttctgaactcactcacaacctctcttaacgctcgtgtgaaagcc 843  
 |||||||  
 Db 849 CCACCTGTGAGAACTTTTGCTAACTACTACAAACCTTTTAAAGCCTTCGTGAGAACCC 908  
 QY 844 atggaccgtatggtaacattaccctctgacggttacccaagccagattcgtctgaac 903  
 |||||||  
 Db 909 ATGGACCGTATGGGTAACTTACCCCTCTGTACGGGTACCCAGGCCAGATTCTGTCTGAC 968  
 QY 904 tgcagagtggtcaacagcaactct 927  
 |||||||  
 Db 969 TGCAGAGTGGTCAACACCACTCT 992  
  
 RESULT 3  
 Q37614  
 ID 037614 standard; DNA; 1072 BP.  
 AC 037614;  
 DT 23-JUN-1993 (first entry)  
 DE Encodes HRP/IGF-II fusion protein.  
 KW Insulin receptor; E.coli periplasmic protease resistance;  
 KW wound healing; osteoporosis; muscle-wasting; growth disorder;  
 KW horseradish peroxidase; inclusion bodies; ds.  
 OS Synthetic.  
 FH Key  
 FT Location/Qualifiers  
 FT cds 11..1060  
 FT /\*tag= a  
 FT /note= "HRP/IGF-II fusion gene"  
 FT misc\_feature 11..856  
 FT /\*tag= b  
 FT /function= HRP coding sequence  
 FT /note= "encodes first 282 amino acids of HRP"  
 FT misc\_feature 857..1060  
 FT /\*tag= c  
 FT /function= IGF-II coding sequence  
 FT WO9303152-A.  
 PN 18-FEB-1993.  
 PD 27-JUL-1992; G01389.  
 PF 29-JUL-1991; GB-016325.  
 PR 05-FEB-1992; GB-002401.  
 PA (BR1-) BRITISH BIO-TECHNOLOGY LTD.  
 PI Bawden L, Edwards RM;  
 DR WPI: 93-076509/09.  
 DR P-PSDB: R32760.  
 PT New insulin-like Growth Factor II analogues - for treating growth  
 PT disorders, osteoporosis and muscle wasting diseases, have less  
 PT affinity for insulin receptors than unmodified IGF-II  
 PS Example 4; Page 44-46; 71pp; English.  
 CC This sequence encodes a fusion protein comprising the first 282  
 CC amino acids of HRP fused to IGF-II via a methionine residue to  
 CC allow for the release of mature IGF-II by cleavage with CNBR. The  
 CC fusion lacks the final 27 amino acids of HRP. This fusion protein  
 CC was not ideal for production of IGF-II; the IGF-II is only 19% of  
 CC the total fusion protein and the HRP sequence contains two internal

CC methionine residues which act as additional cleavage sites.  
 CC The IGF-II is the wild-type sequence, i.e. with Arg residues at  
 CC positions 37 and 38. The invention covers analogues of this sequence  
 CC in which one or both of the Arg residues are subd. by other amino  
 CC acids (pref. non-basic and esp. Gln). The mutants are protease-  
 CC resistant and have reduced affinity for insulin receptors.  
 SQ Sequence 1072 BP; 259 A; 304 C; 255 G; 254 T;  
  
 Query Match 91.0%; Score 843.8; DB 1; Length 1072;  
 Best Local Similarity 99.8%; Pred. No. 2e-264;  
 Matches 845; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
 QY 1 atgcagttaacccctacatcttcaacataagctgcgtcccaagtgccaacatcgttcgc 60  
 |||||||  
 Db 11 ATGCAGTTAACCCCTTACATTTCTACGACATATAGCTGTCCAAACGTCTCCAACTGTTCCG 70  
 QY 61 gacacaatcgtcaagagctcgaatccgaatcccaagatcgctctcaatatactgtctg 120  
 |||||||  
 Db 71 GACCAATCGTCAACGAGCTCAGATCCGATCCAGATCCGCTCTCATATTACGTCTG 130  
 QY 121 cacttcgatgactgcttcgtgaatggttgagcgtctgcatattactggacaacaccac 180  
 |||||||  
 Db 131 CACTTCATGACTGCTTCTGTGAATGTTGCGACCTGACATATTACGTGACAAACACACC 190  
 QY 181 agttccgcaactgaaagatgcatctcggaagcctaaacagcgcaggggcttccagtg 240  
 |||||||  
 Db 191 AGTTCCGCACTGAAGAGATGCAATTGCGGAACCTTAACAGCCGACGGCTTCCAGTG 250  
 QY 241 atcgatcgcataagagctgcggttgatgtagcagcaltgcccagaacagtcagtlgtcagac 300  
 |||||||  
 Db 251 ATCGATCGCATGAAGGCTGCGGTTGAGTCAAGCATGCCACAGCAACAGTGTGACGAC 310  
 QY 301 ctgctgactatagctgctgcgaacagagcgtgactcttcagggcggagccgtctggaagt 360  
 |||||||  
 Db 311 CTGCTGACTATAGCTGGGCAACAGAGCGTGACTCTTCAGCGGACCGTCTGTGAGAGTG 370  
 QY 361 ccgctgctgcaagctgaactccctacagcagcatctctagatctcggcacaagccaaacttgc 420  
 |||||||  
 Db 371 CCGCTCGGTGCACGTGACTGCTCCTACAGGCACTTCTAATCTGGCCAAAGCCCACTTGCT 430  
 QY 421 gctccatcttccaccctgcgcccagctgaagatagctcttgaagaacgtggtgtcaatcgc 480  
 |||||||  
 Db 431 GCTCATTTCTTACCCCTGCCCCAGCTGAAGATAGCTTTGAAACGTTGGGTGAATCGC 490  
 QY 481 tgcagtgacctgtgctcgtcgtccgagagacacacattggaaagaacagtgtagtgc 540  
 |||||||  
 Db 491 TCGAGTGACCTTGTGGCTCTGTCCGGAGGACACACATTTGGAAAGAACCACTGTAGGTTTC 550  
 QY 541 atcaatgataggctctacaatttcaagaacacgtgttacctgaccccaagctgaacact 600  
 |||||||  
 Db 551 ATCAATGATAGGCTCTACAAATTTACGACACCTGTGGTTACTGTCGCCCAAGCTTAACACT 610  
 QY 601 acgtatctccagacactgaagagctgtgcccactgaatgcaacacctcagtgactagtg 660  
 |||||||  
 Db 611 ACGTATCTCCAGACACTGAGAGGCTTGTGCCACTGATGAGCAACCTCAAGTCACTAGTG 670  
 QY 661 gacttgatctgcgagcccaacacatcttcgataaacaagtactatgtgaactagaggag 720  
 |||||||  
 Db 671 GACTTGTATCTGCGGACCCCAACCATCTTGTGAACAAGTACTATGTGAATCTGAGAGAG 730  
 QY 721 cagaaaaagccgtgaatgaagatgatacaagaactgttttagcagtcagcagccagcaaac 780  
 |||||||  
 Db 731 CAGAAAGCCGTGATACAGAGATGATCAAGAACTGTTTGGCACTCAACAGCCATGACACC 790  
 QY 781 atcccaatgtgagaagtttctgaactctactcaaaactcttcttaagccttcgttgaa 840  
 |||||||  
 Db 791 ATCCACTGGGGAAGAAATTTTGTGTAACCTACTCAAACTCTTTAAGCCTTGTGTGAA 850  
 QY 841 gccatgg 847  
 || |||||  
 Db 851 GCTATGG 857

```
RESULT 4
ID N93078 standard; DNA; 927 BP.
AC N93078:
DT 10-MAR-1993 (revised)
DE 25-JUN-1989 (first entry)
DS Sequence encoding horseradish peroxidase (HRP) isoenzyme C.
KM Horseradish peroxidase (HRP) isoenzyme C; detection reagent;
KW superoxide radicals (SOD); ss.
FH Key Location/Qualifiers
FT cds 1..924
FT BP-299682-A. /*tag= a
PD 18-JAN-1989.
PF 07-JUL-1988; 306222.
PR 17-JUL-1987; GB-016938.
PA (AMER-) Amersham Int. PLC.
PI Chiswell DJ, Ortlepp SA;
PI MPI; 89-017184/03.
DR P-PSDB; P94810.
PT New synthetic DNA encoding horseradish peroxidase C isoenzyme or variants
PT or active fragments, useful eg as detection reagent.
PS Disclosure; fig.6; 19pp; English.
CC Horseradish peroxidase isoenzyme C (HRP-C) is produced from this sequence
CC by recombinant techniques. Conservative variant can be encoded by intro-
CC ducing base changes resulting in amino acid (AA) substitutions. Proper-
CC ties of this enzyme are essentially the same as native HRP-C. A pepide
CC conty. at least 1 HRP-C epitope can also be encoded by restriction enzyme
CC cleavage resulting in peptides of different lengths (conty. active AA's
CC Arg-183 and Tyr-185). HRP-C is a scavenger of toxic superoxide radicals
CC (SOD) and is useful as a detection reagent. It provides enhanced illumina-
CC tion by catalysis of a substrate such as luminol, in the presence of a
CC light-output enhancer.
SQ Sequence 927 BP; 247 A; 262 C; 199 G; 219 T;

Query Match 60.7%; Score 563; DB 1; Length 927;
Best Local Similarity 75.6%; Pred. No. 3.4e-173;
Matches 698; Conservative 0; Mismatches 225; Indels 0; Gaps 0;

QY 4 cagttaccctacatctctcgaacaatagctgtcccaagctgtcccaacatcgttcgcgac 63
DB 1 CAGTTAACGCCGCGCTTTTACGATTAACAGCTGTCCAAACGCTTCGATATTGTTCGGCAT 60
QY 64 acatcgtcaacagctcagatccgatcccaagatcgtcttcaatattacgttcgcac 123
DB 61 ACCATGCTGAATAGCTGCGGCTCCGATCCACGTAATTCAGCTAGCATCTTTCGTATCAT 120
QY 124 ttccatgaactcgttgaatgtgttcgcagctagacatattacttggaacaacacacagc 183
DB 121 TTTCATGACTGCTTTGTAACGATGTGACGCGTCATCTCTGGAACATACACAGC 180
QY 184 ttccgcacgtgaaagatgcatcctcggaagcgttaacagcgcagggcgttccagtgatc 243
DB 181 TTCGCAAGTAGAAGAGCGCATTTTCGGGAACGCAATTCGGCCCGGCTTTTCCAGTCATC 240
QY 244 gatcgaataaagctccgtttgagtcagcagtcgccaagacagtcagttgtgaagcctg 303
DB 241 GATCGCATGAAGCTGAGTCGAGTCGGCTGCGCCGAGAACTTTTCGTGTGACAGCTTG 300
QY 304 ctgactaagctcgcgaacagagctgaactctgcagcgcgagcgtctggaagtgacg 363
DB 301 CTGACCATCGCACCCGACGAATGGTCAACCTTAGCGGGCGGCTTCATGGCAGATACCA 360
QY 364 ctgcgtcagcagtgatccctcctacagcattcctgattctggccaagccaaactgtcgtc 423
DB 361 TTGAGCAGCTGCTACCTCCACAGGCTTTCTGATCTACTAATCAATTCATACGGGCG 420
QY 424 ccaatttaccctccccaagcgtgaaggaatagctttagaagaacgtgggttgaaatgctcg 483
DB 421 CCAATTTTACCTCTACCAACATTAAGGATTCAATTTTCGTACCTAGGCTGGAACCGCTCG 480
```

```
QY 484 agtacccttggtcctcgtccgaggaacacacatttggaaagaacagtgtaagttcaltc 543
DB 481 AGTATCTATGATAGCTCTTTCTGTGGCCACACTTTGGAAAGAACAGTGGCCATTCATC 540
QY 544 atggatagcctctacaatttcagaacactcgggtttaccctggaacccagcgtgaactaag 603
DB 541 ATGGACCGAATTTGACAACCTTTTCAACACAGGACTACCGGATCCACGGCTCAATPAGAC 600
QY 604 taccccaagactggagagcgtgtgccactgaatggcaactcagtgcaactagtgagac 663
DB 601 TACTCTACAGACACTCCGGGACTATGCCCCCTTTAAAGGAAACCTAAGCGCGCTAGTCGAC 660
QY 664 ttgatctcgcgaccccaacacalccttcgataaacaagtactatgtgaactagaagagcag 723
DB 661 TTGAGCTTAGCTAGCGCAACAAATCTTCGACAAACAGTACTATGTAATCTAGAGAACAA 720
QY 724 aaagccctgatacagaagtgatacagaactcgtttagcaggtccagaagccactgaccac 783
DB 721 AAAGGACTTATTCATCTGTCAAGAGCTCTTCTCATCAACCGGAATGCAACAGACAGATC 780
QY 784 ccactgtgagaagttttgctactctactcaactctttaaagcccttcgtggaagcc 843
DB 781 CCAGTGTGCGGTAGCTTTGCGAACTCAACACAACTTTTCAATGCAATTCGTAGAGGCC 840
QY 844 atgaccgctatggttaacatlaacccctcgtgacggtgtaaccaaggtcagatcgtcgaac 903
DB 841 ATGATTCGATGATGGGAATATACAGCCACATCAGCGGTACCAAGTCAATCAGATTGAAC 900
QY 904 tgcagagtgtgtaacagcaactc 926
DB 901 TGCCGAGTGTGTAATTCGAACTC 923

RESULT 5
ID N81508 standard; DNA; 924 BP.
AC N81508;
DT 05-NOV-1990 (first entry)
DE Horseradish peroxidase structural gene.
KM Horseradish peroxidase; Armoracia lapathifolia;
KW Armoracia rusticana; ss.
OS Armoracia sp.
FH Key Location/Qualifiers
FT cds 1..924
FT /tag= a
FT /product=horseradish peroxidase

PN J63207386-A.
PD 26-AUG-1988.
PF 19-FEB-1987; 037623.
PR 19-FEB-1987; JP-037623.
PA (TOYM) Toyobo KK.
DR MPI; 88-282535/40.
DR P-PSDB; P81167.
PT New structural gene of horseradish peroxidase -
PT produced under artificially controlled conditions by gene
PT manipulation.
PS Disclosure; P; Japanese.
CC DNA was preferably isolated from tissue cultures of root cells of
CC Armoracia sp. The gene can be inserted into a plasmid for
CC transformation of suitable host cells. The product (HRP) can then
CC be collected from the culture medium.
CC See also N82195.
SQ Sequence 924 BP; 256 A; 224 C; 187 G; 257 T;

Query Match 56.7%; Score 525.6; DB 1; Length 924;
Best Local Similarity 73.1%; Pred. No. 4.8e-161;
Matches 675; Conservative 0; Mismatches 249; Indels 0; Gaps 0;

QY 4 cagttaccctacatctcgaacaatagctgtcccaagctgtcccaacatcgttcgcgac 63
DB 1 CAGTTAACCCCTTACGATTCAGACAAATTCATTAAGTGTCTTAACATTCGACGGGAT 60
```





DT 18-JUL-1990 (first entry)  
DE 3' end half of synthetic horseradish peroxidase gene (HRP)  
KM Horseradish peroxidase; restriction sites; DNA construct;  
KW fusion protein; enzyme assays; immunoassays; synthetic gene.  
OS Horseradish.

FT key Location/Qualifiers  
FT cds 10..937  
FT /\*tag- a

PN MO8903424-A.  
PD 20-APR-1989.  
PF 07-OCT-1988; G00833.  
PR 08-OCT-1987; GB-023662.  
PA (BRRI-) Brit Bio-Tech Ltd.  
PI Edwards RM, Burke JF;  
DR MPI: 89-130042/17.  
DR P-PSDB: P95200.

PT Synthetic gene encoding horseradish peroxidase -  
PT used for efficient prodn. of the enzyme or fusion prods. for  
PT use in enzyme assays, esp. immunoassays

PS Disclosure: :45pp; English.  
CC The 5' half of synthetic HRP is given in n90761; n90760 gives the  
CC full synthetic HRP gene, which contains many restriction sites.  
CC The synthetic gene is prepd. by coupling successive nucleotides  
CC and/or ligating appropriate oligomers. A construct containing bases  
CC 14-937 of the synthetic gene is claimed. The construct can be a  
CC vector where the HRP gene and a desired gene are co-expressed as  
CC a single fusion prod., a single polyclistronic message, or 2 separate  
CC but linked transcriptional units. The fusion proteins are for use  
CC in enzyme assays, esp. immunoassays. Genes carrying mutations that  
CC destroy HRP activity can be used to follow the frequency of  
CC reversion or suppression of the particular mutation introduced  
CC into the gene. The constructs can be used in transgenic animals,  
CC esp. mammals.

SO Sequence 474 BP; 133 A; 124 C; 106 G; 111 T;

Query Match 48.2%; Score 446.4; DB 1; Length 474;  
Best local Similarity 99.8%; Pred. No. 1.9e-135;  
Matches 447; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 480 ctgagtgacctgtgtgctctgtccgagagacacatttggaagaacagtgtagt 539  
DB 9 CTCGAGGACCTTGTGCTGTCCGAGGACACACTTTGGAAAGCAAGCTTAGGTT 68  
QY 540 catcatgagactgactacacatttcaaacactgggttaccgaccccaagctgaacac 599  
DB 69 CATCATGAGTAGGCTCTACAAATTTCAACAACACTGGTTACCTGACCCACGCTGAACAC 128  
QY 600 tacgtatctccagacactgagagctgtgccactgaatggaacactcaagtcaactagt 659  
DB 129 TACGTATCTCCAGACACTGAGAGCTGTGCCACTGAATGGCAACCTCAGTGCACTAGT 188  
QY 660 gagacttgactctcgagaccccaacatcttcgataaactactatgtgatatcagaagga 719  
DB 189 GGACTTTGACTCGGGGACCCCAACCACTTCGATAAACAACTACTATGTGATCTAGAGGA 248  
QY 720 gcgagaaagcgtatatacagagtgatcaagaacgtttagtagtccaaagcccaactgacac 779  
DB 249 GCAGAAAGGCTGATATAGAGTATCAAGAACTGTTTAGAGAGTCCAAAGCCCACTGACAC 308  
QY 780 catcccaactggtgagaaatttgcttaacttaactcaaaccttcttaagcctctgtgga 839  
DB 309 CATCCCACTGGTAGAAGTTTGTCTTAAGTCAAAACCTTCTTAAACCTTCCTGCGGA 368  
QY 840 agccatgagaccgtatagggtacattaccctctgagaggtaccacaagcgcaagttgctc 899  
DB 369 AGCCATGAGCCGTAATGGGTAACTTACCTCTGACGGGTAACCAAGCGAGATTCCTCT 428  
QY 900 gaactcgaagtggtcaacagcaactct 927  
DB 429 GAACGTGAGAGTGTGTCAACAGCAACTCT 456

RESULT 8

ID V30460  
AC V30460; standard; cDNA; 1244 BP.

DT 09-NOV-1998 (first entry)

DE Soybean seed coat peroxidase cDNA.

KW peroxidase; seed coat; soybean; transgenic plant; ss.

OS Glycine max cv. Harosoy 63.

FT key Location/Qualifiers

FT CDS 1..1059

FT sig\_peptide 1..77

FT mat\_peptide 78..1056

FT /\*tag- a

FT /\*tag- b

FT /\*tag- c

PN CA2211018-A.

PD 30-MAR-1998.

PE 19-SEP-1997; 211018.

PR 30-SEP-1996; US-723414.

PR 30-SEP-1996; CA-186833.

PA (MIAC ) CANADA MIN AGRICULTURE.

PI Gilzen M;

DR MPI: 98-388720/34.

DR P-PSDB: W44931.

PT New DNA used in preparation of soybean seed coat peroxidase - used

PT e.g. in diagnostic assays, for de-linking waste paper, for oxidation

PT of primary alcohols, in paper and pulp industries and to encode

PT herbicide resistance in plants

PS Claim 1: Page 42-45; 75pp; English.

CC This cDNA clone includes an open reading frame coding for soybean

CC seed coat peroxidase. A cDNA library constructed from developing

CC seed coat tissue of soybean Epp cultivar Harosoy 63 was screened

CC with a probe corresponding to the conserved active site domain of

CC plant peroxidases to isolate the clone. The structural gene (see

CC V30461) was isolated from a Harosoy 63 genomic library using the

CC cDNA as probe. Vectors, host cells, and transgenic plants are

CC provided, as well as methods for the production of recombinant

CC seed coat peroxidase. Heterologous genes of interest are produced

CC within seed coat cells by propagating a transformed plant with a

CC vector comprising contiguous nucleotides 1-1532 of the genomic DNA.

CC Soybean coat peroxidases are useful industrially, e.g. their high

CC thermal stability even at extremes of pH make them useful in

CC diagnostic assays as an enzyme label for e.g. antigens, and within

CC staining techniques. They are also useful for de-linking printed

CC waste paper, biocatalytic oxidation of primary alcohols, as a

CC replacement for chlorine in the pulp and paper industry or as a

CC formaldehyde replacement for use in adhesives. The peroxidase may

CC also be expressed in an organ- or tissue-specific manner within

CC plants by employing appropriate promoters. Heterologous genes may

CC encode herbicide resistance, biological control of insects/pathogens

CC or proteins of commercial interest (e.g. pharmaceuticals) or that

CC alter e.g. the nutritive value of seeds. The cDNA/genomic DNA are

CC also useful to produce probes to detect polymorphisms that

CC distinguish Epp and eep genotypes; varieties containing a

CC dominant Epp allele accumulate large amounts of peroxidase in the

CC hourglass cells of the subepidermis, whilst homozygous recessive

CC eep genotypes do not accumulate peroxidase in the hourglass cells

CC and have much reduced seed coat peroxidase activity.

SO Sequence 1244 BP; 358 A; 261 C; 250 G; 375 T;

Query Match 25.9%; Score 239.8; DB 1; Length 1244;  
Best local Similarity 55.3%; Pred. No. 4.4e-68;  
Matches 510; Conservative 0; Mismatches 407; Indels 6; Gaps 2;

QY 4 cagttaccctacatcttaagacataagctgcccaagtgctccaaatcgctgcgac 63  
DB 79 CAGCTTACTCTCTACGTTCAGAGAAACATGTCCAAATCTGTCCCTATTGTTTGA 138  
QY 64 acatcgtcaagcaggtccagatccgagatcgctgcttcattatcgtctgcac 123  
DB 139 GTATCTTCATGCTCTTTTCACCGATCCCGAATGCGGCGCAGTCTCATGAGGCTTCA 198

QY	124	ttccatgactctcttcgtagaatggtctgtggaagctagcactatctatctggaacacacaccacag	183
Db	139	TTTCATGATTCCTTTGTTCAAGGTTTGATGATCATCACTTTTGTGAAACAACCTGATACA	258
QY	184	ttccgcactggaagaagatgacatctcggggaacgctaacagcgccgaaggccttcacgtgac	243
Db	259	ATAGAAACCGAGCAAGATGAGCTTCCAAATATCAACTCATTAAGAGGATTGGACGTTGTG	318
QY	244	gacgcataagagagctctccgtttgactgacgacatgcccacgaaacagtcagttgtgcagactg	303
Db	319	AATGACATCAAGACAGCGCGTGGAAAAATAGTTGTCTCAGACACACAGTTTCTGTGCGATATT	378
QY	304	ctgactataagctctgccaacagaagcgtaactctctgcagcggaagccgtctccgtagaagtcgcg	353
Db	379	CTTGCTATTGGACGTGGAATATGCTTCTGTTCTTGCGAGAGAGTCCAGGATGGCCAGTTCCA	438
QY	364	ctcgctgaagctgactccctcctacaggaatccctagatctggtccaaagcgaacttgcctgct	423
Db	439	TTAGGAGAGAGGACAGCTTATACAGCAACCGAACCTTGGCAATTCAAAACCTTCCAGCA	498
QY	424	ccatctcttcaccccttgcgcccaagctgaagatagactctttagaagacgttggcttgaatcgctg	483
Db	499	CCTTTCTTCAACGTCACACTCAACTTAAGCTCTCTTGTCTGTTCAGAGTCTCAAA--CAC	555
QY	484	agtgacactgtgctctgctctccggaaggaacacacatctggaaagaacagtcagtgatctac	543
Db	556	CTTGATTATTAGTTTACACTCTCAGTGGTGCATACGTTTGGAGAGAGCTCGTGCAGATCATTC	615
QY	544	atggaatagcctcctaacttttagaanaacactgggttaccttgaccccaacgctggaacactag	603
Db	616	ATTAACCGATTTATCAACTTCTAGACAACTGGAAACCTTATTCACACTGTGAACACACACA	675
QY	604	tatctccagaacacgtagagagctgtgcccactgtagatgcaacactcagtgcaactagtgac	663
Db	676	TACTTGAAGATATTGGCGTGCAGAAATGGCCCGCAGATGCACTGGGGATTAACCTCACCAAT	735
QY	664	cttgatctcgggaaccccaacatcttcgataacaagaactatgtgaactctagaaggagcag	723
Db	736	TTGGACCTGAGACACACCTGATCAATTTTGACAAAGATATTAATCCAACTTCTGACAGCTC	795
QY	724	aaagccctgatacagagatgatacgaagaactgtttagcagctcagacgcacatgacacatc	783
Db	796	AATGGCTTACTTCAAGAGACCAAGAACTTTCTCCACTCCTGGTG--CTGATACCATTT	852
QY	784	ccacgtgtagaagaagctttgtcctaactctactcaaacctctttaaagcctctcgtggaagcc	843
Db	853	CCCATTTGTCATTAAGCTTACAGCAGTACAGATTAATCTTTCTTTCCAACTTTAGAGTTTCA	912
QY	844	atggaacgcatlgyglatcaatctaccctcttgacgaggtlaaccaaggccagatctgctgaac	903
Db	913	ATGATAAAAAATGGGTAATATTGGAAGTCTGACCTGGGGATGAAGAGAAATTTCCGTTGCAA	972
QY	904	tgcagagtgctgacacagcaactc 926	
Db	973	TGTAAATTTTGTGAATGAGAGCTC 995	
RESULT 9			
ID	V59077	standard: cDNA: 1244 BP.	
AC	V59077:		
DE	02-FEB-1999	(first entry)	
DR	Soybean seed coat peroxidase cDNA.		
KW	Peroxidase: seed coat; soybean; transgenic plant; ss.		
OS	Glycine max cv. Harosoy 63.		
FN	key	Location/Qualifiers	
FT	CDS	1..1059	
FT	FT	/*tag= a	
FT	FT	1..78	
FT	FT	/*tag= b	
FT	FT	79..1056	
FT	FT	mat_peptide	
FT	FT	primer_bind	
FT	FT	complement (53..68)	

FT		/tag= d	
FT		/note= "primer prx9+"	
FT	primer_bind	complement (143. .162)	
FT		/tag= e	
FT		/note= "primer prx12+"	
FT	primer_bind	254. .273	
FT		/tag= f	
FT		/note= "primer prx10-"	
FT	primer_bind	complement (280. .299)	
FT		/tag= g	
FT		/note= "primer prx2+"	
FT	primer_bind	885. .904	
FT		/tag= h	
FT		/note= "primer prx6-"	
PN	CA2186833-A.		
PD	31-MAR-1998.		
PE	30-SEP-1996; 186833.		
PR	30-SEP-1996; CA-186833.		
PA	(MTAC ) CANADA MIN AGRICULTURE.		
PI	G1jzen M;		
DR	WPI: 98-521791/45.		
DR	P-PSDB: W73138.		
PT	New DNA encoding soya seed-coat peroxidase and the promoter - useful		
PT	for expression of heterologous genes in the seed coat, which may		
PT	give herbicide resistance, viral protection and pest and pathogen		
PT	control, and may improve the taste or nutritional value of soya		
PT	beans.		
PS	Claim 1: Fig 1: 61pp: English.		
CC	This nucleotide sequence includes an open reading frame coding for		
CC	a seed coat peroxidase (see W73139) of soybean. To isolate the		
CC	peroxidase transcript, a cDNA library was constructed from		
CC	developing seed coat tissue of the EBP cultivar Harosoy 63, and		
CC	was screened with a 17-mer oligonucleotide probe (see V59079)		
CC	corresponding to the conserved active site domain of plant		
CC	peroxidases. 6 Clones contained identical cDNA transcripts. A		
CC	genomic clone (see V59078) was also isolated and shown to contain 4		
CC	exons. The 5' regulatory region of the gene is also claimed. The		
CC	heterologous expression, as well as vectors and host cells to be		
CC	used for the expression of the seed coat peroxidase, are disclosed.		
CC	Probes derived from the cDNA or genomic DNA can be used to detect		
CC	polymorphisms that distinguish Ebp and eep genotypes. The		
CC	seed-specific regulatory region can be used to control expression		
CC	of genes of interest within the seed coat of transgenic plants.		
SQ	Sequence 1244 BP; 358 A; 261 C; 250 G; 375 T;		
<hr/>			
Query Match	25.9%;	Score 239.8;	DB 1; Length 1244;
Best Local Similarity	55.3%;	Pred. No. 4.4e-68;	
Matches 510; Conservative	0;	Mismatches 407;	Indels 6; Gaps 2;
OY	4	cagtaaacccacagtcttaccagaatagtcgtcccaacgftccaaacgttcgcgaac	63
DB	79	CAGCTTACTCCTACGTTCTTACAGAGAACAACATGTCCAATACTTCCTATTGTGGTA	138
OY	64	acaactgcgaacagagctcagatcgcatccagatgcgtcgtccaataltaacgttcgac	123
DB	139	GTAATCTTGCGATGCCTTCCTTTCACGCCATGCCCAATCGGGCCAGTCAATGAGGCTTCA	198
OY	124	tccatgacctcttcgtgaaatggttgcgagcgttagcataattactcygacaacaccagct	183
DB	199	TTCATGTGATTGCTTGTTGTTCAAGGTGTGATGATGATTTGCTGAACAACTAATACA	258
OY	164	ttccgactgaaaagatgcattcgctgggaacgctaacaagccagaggccttcacgtgatic	243
DB	259	ATAGAAAGCGACCAAGATGCACTTCCAAATATCAACTCATATAGAGGATGGACCTTGC	318
OY	244	gatccacgaagaagctgcgcttgatcgaagalgtcccaagaaacgaatgltcacacctg	303
DB	319	AATGACATCAAGACACGCGGTGGAANAATAGTTGTCACAGACACAGTTTCTTGTCATAT	378
OY	304	ctgactatagctgcgaacagaagcgtgacctcttcgagcggaacgccttcctygaaagtgcg	363
DB	379	CTTGCTATTGCAAGTAAATAGCTTCTGTTGCGGAGGAGTCCAGATNGCGCACTTCCA	438

```
QY 364 ctgcgtgcagctgactccctacaggaatcctcctgactcgtcccaagccactgctgct 423
    || || || || || || || || || || || || || || || || || || || || ||
D 439 TTAGGAAGAGGAGGACAGCTTAACAGCAAAACCTTGCATTAACAACTCCAGCA 498
QY 424 cactctcccccctgcccagctgaaggaatgactttagaagcgtgggtcgtgactgctg 483
    || || || || || || || || || || || || || || || || || || || || ||
D 499 CTTTCTCAACCTCACTCACTTAAGGCTCTTCTGCTTTCAGAGCTCCAA--CACC 555
QY 484 agtgaacctgagctcgtccgagagacacatttggaagaaccagtgtagtctatc 543
    || || || || || || || || || || || || || || || || || || || || ||
D 556 CTGATTGATTACACTCTCAAGTGCTATGCTTGGAGAGCTCGGTGCATGACTTC 615
QY 544 atggataagctctacaattcagaacaactgggttaccctgaccccaagctgaaactag 603
    || || || || || || || || || || || || || || || || || || || || ||
D 616 ATAAACCGATTATTAACAACCTTACAGCAACACTGAAACCCCTGATCCAACTGCAACACACA 675
QY 604 taltccagaactgagagagctgtgcccactgaaatgagcaactgaatgactagtgag 663
    || || || || || || || || || || || || || || || || || || || || ||
D 676 TACTTGAAGATTATGCTGCAAGATGCCCCCAAGATGCAACTGGGGATTAACCTCACCACAT 735
QY 664 ttgatctgcgagccccaacatctcgtatacaagaactgtgaatctagaagagag 723
    || || || || || || || || || || || || || || || || || || || || ||
D 736 TTGGACCTGAGCACACCTGATTCATTTGACACAGATAGTACTCCAACTCTTCTGCACTC 795
QY 724 aagagctgatacagagtgatcaagaactgtttagcagtcgacagccactgacacatc 783
    || || || || || || || || || || || || || || || || || || || || ||
D 796 AATGCGCTTACTTCAAGAGTACCAAGAACTTTCTCCACTCTCGTG--CTGATACCATTT 852
QY 784 cccctgtgaggaagtttctgtaactctactcaaacctctctttagcgtctgtgaaagc 843
    || || || || || || || || || || || || || || || || || || || || ||
D 853 CCATATGTCATATAGCTTACAGCACTTACCAAGATTAATCTTTTCCAACTTGAAGTTTCA 912
QY 844 atgacccgtatggttaacatctaccctctgacggtgtaacccaagccagatctgtgaac 903
    || || || || || || || || || || || || || || || || || || || || ||
D 913 ATGATATAAATGSGTATATTTGAGAGTGCTGACTGCGGATGAAGAGAAATTCGCTTGCAA 972
QY 904 tgcagaagtggtcaacagcaactc 926
    || || || || || || || || || || || || || || || || || || || || ||
D 973 TGTAAATTTGTGAATGAGACTC 995

RESULT 10
V63720
ID V63720 standard; cDNA; 1355 BP.
AC V63720;
DE 12-APR-1999 (first entry)
KW Maize per5 root preferential cationic peroxidase cDNA.
KW Peroxidase; per5 gene; maize; corn; transgenic plant; promoter;
KW root; ss.
OS Zea mays.
FH Key
FT CDS Location/Qualifiers
FT 53..1054
FT /*tag= a
FT sig_peptide 53..148
FT /*tag= b
FT mat_peptide 149..1051
FT /*tag= c
FT polyA_signal 1293..1298
FT /*tag= d
PN WO9856921-A1.
PD 17-DEC-1998.
PE 10-JUN-1998; 011921.
PR 12-JUN-1997; US-049752.
PA (DOWC) DOW AGRSCIENCES LLC.
PI Ahnley M, Armstrong K, Belmar S, Folkerts O, Hopkins N,
PI Menke MA, Paredy D, Petolino JF, Smith K, Woosley A,
DR P-PSDB; W87893.
DR New isolated regulatory sequences for transgenic plants - which are
PT derived from the maize root preferential cationic peroxidase protein
PS (per5) gene
Example 3; Page 91; 150pp; English.
```

CC This is the nucleotide sequence of a maize per5 cDNA clone encoding  
CC root preferential cationic peroxidase (see W87893). The clone was  
CC isolated by screening maize root, stem, leaf, kernel and tassels RNA  
CC using probes (see V63718-19) based on tobacco peroxidase conserved  
CC domains. A probe generated from per5 cDNA was used to isolate the  
CC per5 gene (see V63717). Regulatory sequences of the per5 gene,  
CC including the promoter, introns and 3' untranslated region (3'UTR),  
CC are used in claimed recombinant gene cassettes for controlling  
CC expression of recombinant genes in selected tissue, especially the  
CC root, of transformed plants, particularly maize.

Sequence 1355 BP; 333 A; 344 C; 321 G; 351 T; 6 U;

Query Match 24.9%; Score 230.4; DB 1; Length 1355;

Best Local Similarity 54.8%; Pred. No. 5; 2e-65;

Matches 503; Conservative 0; Mismatches 406; Indels 9; Gaps 2;

```
QY 7 ttaaccctacattctacgaacaatagctgcccaagctgtccaacatcgltcgagacaa 66
    || || || || || || || || || || || || || || || || || || || || ||
D 140 TTGTTCCACAGTTCATATGCAATGCTGCTGCCAAGCGAAGAGATTGTGCACTCAATT 199
QY 67 atcgtaacagagctcagatccgattccagatcgtcgtctcaattatagctgcacactc 126
    || || || || || || || || || || || || || || || || || || || || ||
D 200 GTGGCACAGGCTGTGGCCAAAGAGACAGATGGCGCATTTTACTGACACTGCAATTTTC 259
QY 127 catgactgtcttgtaatggtltgcagcgttagcatactactgagacaacaccagattc 186
    || || || || || || || || || || || || || || || || || || || || ||
D 260 CANGACTGCTTTTTCAAAGGCTCGCATGCTGCTGCTGTTGACAAACAGACACACATTA 319
QY 187 cgcactgaaagagatgaltcaltcggaagcgttaacagcgcgaagggcttccagtgatcat 246
    || || || || || || || || || || || || || || || || || || || || ||
D 320 GTTAGTAGAAGAGGTCACCAACCCAGAACAGAACTCCCTCAAGGGGCTTGGAGTGCAGAC 379
QY 247 cgcataagagctgcgtgtgtagtgcagatgccagagacagtcagtgatgagagccgctg 306
    || || || || || || || || || || || || || || || || || || || || ||
D 380 CAAATTAAGGCTCTCTTGAAGCTGCTGCTGCCAGGACAACTCTCTGTCGCGCATTTGTT 439
QY 307 actatagctgcgcaacagagcgtgactctgcagagcgagacgctcgtgagagtgacgctc 366
    || || || || || || || || || || || || || || || || || || || || ||
D 440 GCCCTTGGCGGCTGCTGATTTCCACCGCCCTGGTGGTGGAGACATATGCGGACGTGCACATT 499
QY 367 ggtcgaagtgactccctacagagcattcctagatctgagccaaagcgaactgctgtcca 426
    || || || || || || || || || || || || || || || || || || || || ||
D 500 GGCCGAGAGACTCGCTCGGTGCAACATCCAGGCTCCCAACATCAATCCAGGCCCC 559
QY 427 tcttaccctgcccagcgtgaaagatagctttagaagaagtggtgctgtagctcgtgagt 486
    || || || || || || || || || || || || || || || || || || || || ||
D 560 AACAAACACACTCCCTCACTATCATACCAAGTTCAAGCGCCAGGCGCTCAATGTTG--TT 616
QY 487 gacctgtgctctgtcgcgagagacacacatttggaagaacaggtgtagtcatcatcgt 546
    || || || || || || || || || || || || || || || || || || || || ||
D 617 GATGTTGTGGCCCTCTTCAGAGTGTCACACCATTTGATGTCGTCGTCATGATTTCCGG 676
QY 547 gataagcttacaattcagaacacactggttgaactgagcccaagcgtgaacactagat 606
    || || || || || || || || || || || || || || || || || || || || ||
D 677 CAGAGGCTATACAAACAGACAGGCAATGGCATGCGTAGCAGCACTGATGATGATCTAC 736
QY 607 ctccagacactgagagagctgtgtgcccactgaaatgagaaactgagcactagtagagctt 666
    || || || || || || || || || || || || || || || || || || || || ||
D 737 GCCGCAAGCTGAGGAGGAGATGCCCGCTGTGGTGAGCAACAACTTCCCTTG 796
QY 667 gatctgcgagcccaacacatctcgaatacaagtaactatgtaactagaagagcaaaa 726
    || || || || || || || || || || || || || || || || || || || || ||
D 797 GACTTCATCACCCTCGCCCAAGTTTGACAAATTTTACTACAGAACTCTCGCGGCGCAAG 856
QY 727 ggcctgtatacagagtgatcaagaactgtttagcagtcgagcgcacgtacacatccca 786
    || || || || || || || || || || || || || || || || || || || || ||
D 857 GGCCTTCAAGCTCTGATGATGATTCTGTAAACCA-----AGAGCGCTCAGACAGCGGCC 910
QY 787 ctggtgagaagtttgcctactactactcaaacctctttaaagcctgtgtagaagcagtg 846
    || || || || || || || || || || || || || || || || || || || || ||
D 911 CTCGTGAAGGCATATGCTGCTGATGTCAATCTCTTCCAGACACTTTCAGACAGTCTATG 970
```







PF 27-JUL-1992; G01389.  
 PR 29-JUL-1991; GB-016325.  
 PR 05-FEB-1992; GB-002401.  
 PA (BRBI-) BRITISH BIO-TECHNOLOGY LTD.  
 PI Bawden L, Edwards RM;  
 DR WPI: 93-076509/09.  
 DR P-PSDB; R32761.  
 PT New insulin-like Growth Factor II analogues - for treating growth  
 disorders, osteoporosis and muscle wasting diseases, have less  
 PT affinity for insulin receptors than unmodified IGF-II  
 PS Example 4: Page 50; 71pp: English.  
 CC This sequence encodes a fusion protein comprising the first 53  
 CC amino acids of HRP fused to IGF-II via a unique methionine residue  
 CC to allow for the release of mature IGF-II by cleavage with CNBr.  
 CC The fusion sequence is better for expression of IGF-II than a  
 CC fusion of the first 282 amino acids of HRP with IGF-II (see 037614)  
 CC in which IGF-II is only 19% of the total fusion protein and the HRP  
 CC sequence contains two internal methionine residues which act as  
 CC additional cleavage sites. The IGF-II is the wild-type sequence,  
 CC i.e. with Arg residues at positions 37 and 38. The invention covers  
 CC analogues of this sequence in which one or both of the Arg residues  
 CC are substid. by other amino acids (pref. non-basic and esp. Gln).  
 CC The mutants are protease-resistant and have reduced affinity for  
 CC insulin receptors.  
 SQ Sequence 388 BP; 85 A; 113 C; 90 G; 100 T;

## Query Match

Best Local Similarity 17.2%; Score 159.4; DB 1; Length 388;  
 Matches 160; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atgcagltaacccctacatctctacgacacatagctgtcccaagctgtccaaatcgtcgc 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 11 ATGCAGTTAAACCCCTACATTTCTACGACAAATAGCTGTCCCAACGTTCACATCTCGC 70  
 QY 61 gaccacaatgcacagagctcacaatccagatccagatcgctgttcaatattagctgtg 120  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 71 GACACAAATCGTCAAGAGCTCAGATCCGATCCAGATCGCTGCTCAATATTACGTCTG 130  
 QY 121 cacttccatgactgtcgtgaatggtgtgcagcctagcat 161  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 131 CACTTCATGACTGCTCTCGTAATGTTGGACCGCTAGTAT 171

Search completed: October 4, 2000, 06:40:45  
 Job time: 12254 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2000, 05:49:03 ; Search time 74.16 Seconds  
(without alignments)  
1890.518 Million cell updates/sec

Title: US-09-246-451-16

Perfect score: 927

Sequence: 1 atgcagtaacctacatc.....gagtgtcacagcaactct 927

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 262060 seqs, 75620727 residues

Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgnl\_7/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgnl\_7/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgnl\_7/ptodata/1/ina/5C\_COMB.seq:\*  
4: /cgnl\_7/ptodata/1/ina/5D\_COMB.seq:\*  
5: /cgnl\_7/ptodata/1/ina/6\_COMB.seq:\*  
6: /cgnl\_7/ptodata/1/ina/PCTUS\_COMB.seq:\*  
7: /cgnl\_7/ptodata/1/ina/backfilist1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	843.8	91.0	1072	2	US-08-190-029A-9
2	843.8	91.0	1072	3	US-08-462-695-9
3	163.4	17.6	1270	1	US-08-378-588-23
4	163.4	17.6	1270	3	US-08-811-094-23
5	159.4	17.2	1270	6	PCT-US94-11121-23
6	159.4	17.2	388	2	US-08-190-029A-11
7	159.4	17.2	388	3	US-08-462-695-11
8	148.6	16.0	1131	1	US-08-181-271A-22
9	148.6	16.0	1131	1	US-08-449-315-22
10	148.6	16.0	1131	1	US-08-444-803-22
11	148.6	16.0	1131	1	US-08-449-043-22
12	148.6	16.0	1131	2	US-08-456-265A-22
13	148.6	16.0	1131	2	US-08-455-416-22
14	148.6	16.0	1131	2	US-08-455-244-22
15	148.6	16.0	1131	2	US-08-454-876-22
16	148.6	16.0	1131	3	US-08-457-364-22
17	148.6	16.0	1131	3	US-08-456-262-22
18	148.6	16.0	1131	3	US-08-456-240-22
19	148.6	16.0	1131	3	US-08-455-736-22
20	148.6	16.0	1131	4	US-08-971-217-22
21	146	15.7	1181	1	US-08-181-271A-27
22	146	15.7	1181	1	US-08-449-315-27
23	146	15.7	1181	1	US-08-444-803-27
24	146	15.7	1181	1	US-08-449-043-27
25	146	15.7	1181	2	US-08-456-265A-27
26	146	15.7	1181	2	US-08-455-416-27

27	146	15.7	1181	2	US-08-455-244-27	Sequence 27, Appl
28	146	15.7	1181	2	US-08-454-876-27	Sequence 27, Appl
29	146	15.7	1181	3	US-08-457-364-27	Sequence 27, Appl
30	146	15.7	1181	3	US-08-456-262-27	Sequence 27, Appl
31	146	15.7	1181	3	US-08-456-240-27	Sequence 27, Appl
32	146	15.7	1181	3	US-08-455-736-27	Sequence 27, Appl
33	146	15.7	1181	4	US-08-971-217-27	Sequence 27, Appl
34	129	13.9	1064	1	US-08-378-588-15	Sequence 15, Appl
35	129	13.9	1064	3	US-08-811-094-15	Sequence 15, Appl
36	129	13.9	1064	6	PCT-US94-11121-15	Sequence 15, Appl
37	119.8	12.9	1326	3	US-08-671-320-12	Sequence 12, Appl
38	119.8	12.9	1326	3	US-08-868-577-12	Sequence 12, Appl
39	107	11.5	1167	3	US-08-671-320-16	Sequence 16, Appl
40	107	11.5	1167	3	US-08-868-577-16	Sequence 16, Appl
41	105.4	11.4	1314	3	US-08-868-577-10	Sequence 10, Appl
42	105.4	11.4	1315	3	US-08-671-320-10	Sequence 10, Appl
43	98.2	10.6	1191	3	US-08-671-320-14	Sequence 14, Appl
44	98.2	10.6	1191	3	US-08-868-577-14	Sequence 14, Appl
45	79.8	8.6	686	1	US-08-181-271A-28	Sequence 28, Appl

## ALIGNMENTS

RESULT 1  
US-08-190-029A-9  
; Sequence 9, Application US/08190029A  
; Patent No. 5736363  
; GENERAL INFORMATION:  
; APPLICANT: EDWARDS, Richard Mark  
; APPLICANT: BARDEN, Lindsey  
; TITLE OF INVENTION: IGF-II ANALOGUES  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: ALBERETTI & WITCOFF, LTD.  
; STREET: 10 S. WACKER DRIVE, SUITE 3000  
; CITY: CHICAGO  
; STATE: ILLINOIS  
; COUNTRY: U.S.A.  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/190,029A  
; FILING DATE: 28-FEB-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/GB92/01389  
; FILING DATE: 27-JUL-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9202401.7  
; FILING DATE: 05-FEB-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9116325.3  
; FILING DATE: 29-JUL-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: JOHN J. McDONNELL  
; REGISTRATION NUMBER: 26,949  
; REFERENCE/DOCKET NUMBER: 94,062  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-715-1000  
; TELEFAX: 312-715-1234  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1072 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; ORIGINAL SOURCE:

```
ORGANISM: Homo sapiens
FEATURE:
  NAME/KEY: CDS
  LOCATION: 11..1057
  OTHER INFORMATION: /codon_start= 11
  OTHER INFORMATION: /function= "Gene for IGF-II/HRP fusion protein"
  OTHER INFORMATION: /product= "Synthetic DNA"
FEATURE:
  NAME/KEY: misc_feature
  LOCATION: 1..6
  OTHER INFORMATION: /function= "HindIII cleavage site"
FEATURE:
  NAME/KEY: misc_feature
  LOCATION: 1067..1072
  OTHER INFORMATION: /function= "EcoRI cleavage site"
FEATURE:
  NAME/KEY: misc_feature
  LOCATION: 11..856
  OTHER INFORMATION: /function= "HRP coding sequence"
FEATURE:
  NAME/KEY: misc_feature
  LOCATION: 857..1057
  OTHER INFORMATION: /function= "IGF-II coding sequence"
US-08-190-029A-9
```

Query Match 91.0%; Score 843.8; DB 2; Length 1072;

Best Local Similarity 99.8%; Pred. No. 4.9e-277; Matches 845; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
QY 1 atgcagtlaccctcatatctacagcaatagctgtcccaagctgtcccaacatcgttcgc 60
   |||||||
DB 11 atgcagtlaccctcatatctacagcaatagctgtcccaagctgtcccaacatcgttcgc 70
   |||||||
QY 61 gacacaatcgtlcaagagctcaagatcccaagatcgtcttcaatattacgtctg 120
   |||||||
DB 71 gacacaatcgtlcaagagctcaagatcccaagatcgtcttcaatattacgtctg 130
   |||||||
QY 121 cacttccatagctgtctgtgaatgtgttgagagctagcattacttgcgcaaccacc 180
   |||||||
DB 131 cacttccatagctgtctgtgaatgtgttgagagctagcattacttgcgcaaccacc 190
   |||||||
QY 181 agttccgcaactgaagaagatgcatcgcggaagcctaacagcgcagggccttccagtg 240
   |||||||
DB 191 agttccgcaactgaagaagatgcatcgcggaagcctaacagcgcagggccttccagtg 250
   |||||||
QY 241 atcagatcgatgaagcgtcgcgttgagtcagcaltgcccaagaaagtcagttgtcagac 300
   |||||||
DB 251 atcagatcgatgaagcgtcgcgttgagtcagcaltgcccaagaaagtcagttgtcagac 310
   |||||||
QY 301 ctgtctactatagctgcgcaacagagcgtgactcttcgcagggcgagaccgttctgagagtg 360
   |||||||
DB 311 ctgtctactatagctgcgcaacagagcgtgactcttcgcagggcgagaccgttctgagagtg 370
   |||||||
QY 361 ccgctcgctcgaagcttactccctacagagcaltcctagatctggcccaagcgaacttgact 420
   |||||||
DB 371 ccgctcgctcgaagcttactccctacagagcaltcctagatctggcccaagcgaacttgact 430
   |||||||
QY 421 gctcgaatcttcaaccctgcgcccagctgaagatagctttagaagacgttggtctgaaatcgc 480
   |||||||
DB 431 gctcgaatcttcaaccctgcgcccagctgaagatagctttagaagacgttggtctgaaatcgc 490
   |||||||
QY 481 tggagtgactctgtgtctgtctgcggaagacacacatttggaaagacacgtttagttc 540
   |||||||
DB 491 tggagtgactctgtgtctgtctgcggaagacacacatttggaaagacacgtttagttc 550
   |||||||
QY 541 atcaatgataagcttcacaatttcagacaacactgggttlaactgaaccacagctgaaact 600
   |||||||
DB 551 atcaatgataagcttcacaatttcagacaacactgggttlaactgaaccacagctgaaact 610
   |||||||
QY 601 acgtatctccagacacttgagaggtctgtgcccaactgaatggaacactcagtgactagtg 660
   |||||||
```

```
DB 611 ACCTATCTCCAGACACTGAGAGGCTTGTCGCCCACTGAATGCAACCTCAGTGCACATG 670
QY 661 gacttgatctgaggagcccaaccatcttgataaacaagtaagtgtgaatcagaaggag 720
   |||||||
DB 671 gacttgatctgaggagcccaaccatcttgataaacaagtaagtgtgaatcagaaggag 730
   |||||||
QY 721 cagaagggcctgtatagagatgtatcaagaactgtttagcagctcagaagcgaactgacac 780
   |||||||
DB 731 CAGAAAGGCGGTGAGAGAGTTTGCTACTACTCAAACTTCTTAAAGCGCCTTGAGAA 790
   |||||||
QY 781 atcccaatgtgtgagaagtttctgtaacttaactcaactccttcaagccttgaggaa 840
   |||||||
DB 791 ATCCCACTGGTGAGAGTTTGCTACTACTCAAACTTCTTAAAGCGCCTTGAGAA 850
   |||||||
QY 841 gcatgag 847
   |||||||
DB 851 GCTATGG 857
   |||||||

RESULT 2
US-08-462-695-9
; Sequence 9, Application US/08462695
; Patent No. 5854025
; GENERAL INFORMATION:
; APPLICANT: EDWARDS, Richard Mark
; APPLICANT: BARDEN, Lindsey
; TITLE OF INVENTION: IGF-II ANALOGUES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BANNER & ALLEGRETTI, LTD.
; STREET: 10 S. WACKER DRIVE, SUITE 3000
; CITY: CHICAGO
; STATE: ILLINOIS
; COUNTRY: U.S.A.
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,695
; FILING DATE: 5-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/190,029
; FILING DATE: 28-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB92/01389
; FILING DATE: 27-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9202401.7
; FILING DATE: 05-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9116325.3
; FILING DATE: 29-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: JOHN J. McDONNELL
; REGISTRATION NUMBER: 26,949
; REFERENCE/DOCKET NUMBER: 94,062-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-715-1000
; TELEFAX: 312-715-1234
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1072 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; FEATURE:
```

```

; NAME/KEY: CDS
; LOCATION: 11..1057
; OTHER INFORMATION: /codon_start= 11
; OTHER INFORMATION: /function= "Gene for IGF-II/HRP fusion protein"
; OTHER INFORMATION: /product= "Synthetic DNA"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..6
; OTHER INFORMATION: /function= "HindIII cleavage site"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1067..1072
; OTHER INFORMATION: /function= "EcoRI cleavage site"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11..856
; OTHER INFORMATION: /function= "HRP coding sequence"
; OTHER INFORMATION: (fragment)"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 857..1057
; OTHER INFORMATION: /function= "IGF-II coding sequence"
US-08-462-695-9
```

```

Query Match          91.0%; Score 843.8; DB 3; Length 1072;
Best Local Similarity 99.8%; Pred. No. 4.9e-277;
Matches 845; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```

QY 1 atcgagtaacccctacatctcagacaatagctgtcccaagctgtccaaacatcgttcgc 60
    |||||||
DB 11 atgcagttacccctacatctcagacaatagctgtcccaagctgtccaaacatcgttcgc 70
    |||||||
QY 61 gacacaaatcgtcaagagcctcagatccgatcccaagagctcgttcaatataatcgtctg 120
    |||||||
DB 71 GACACAAATCGTCAACGAGCTCAGATCCGATCCGAGATCCGCTGTTCAATTAACGTCTG 130
    |||||||
QY 121 caactcctatcagctcgtcgaaatgtgtcgacagctagacatatactcggacaacaccac 180
    |||||||
DB 131 CACTTCATACACTGCTCTCGTAATGCTGGGAGCCCTAGCATATTACTGGACAACACCACC 190
    |||||||
QY 181 agtttcgcacgtcaaaagagatcgtcggagacgtcaacagcgccagggcgttcacgtg 240
    |||||||
DB 191 AGTTTCGCGACTGAAAAGATGATTCGGGAGACGCTTAACGCGCAAGGGCTTTCAGTGT 250
    |||||||
QY 241 atcgatcgatgaagagctgcgttgatcagcatgcccagaagaacgtcagttgtgcagac 300
    |||||||
DB 251 ATCGATCGATGAAGGCTGCGGTTGAGTCAAGATGCCACGAACAGTCAGTTGTGCAGAC 310
    |||||||
QY 301 ctgctgactatactcgcgaacagagcgttactcttcagcgcgagccgtcctgtgaagtg 360
    |||||||
DB 311 CTGCTGACTATAGCTGGGCAACAGAGCGTGACTCTTTCAGCGGAGCGCTCTGTGAGAGTG 370
    |||||||
QY 361 ccgctcggtcgaagtgaccctcctaaggaatccctagatctgcccacgcgaacttgct 420
    |||||||
DB 371 CCGCTCGGTGAGAGTGAATCTCTTACAGGCACTTCTAGATCTGGCCCAACGCAACTTGCT 430
    |||||||
QY 421 gctccatcttcgaccccgcccgagctgaagatagctttagaaacgtgggtctgaatcgc 480
    |||||||
DB 431 GCTCCATTTCTTCACCCCTGCCCAAGCTGAAGATAGCTTTGAAMAAGTGGGTCTGAATCGC 490
    |||||||
QY 481 tcgaagtaactctgtgctcgtcgtccgagagacacacatttgaagaagaccagtgagttc 540
    |||||||
DB 491 TCGAGTGAAGCTGTGCTGCTCTCGGAGGACACACATTTTGAAGAAGACCACTAGAGTTTC 550
    |||||||
QY 541 atcaatgatataggtctacaatttcagcaacacttggttacttgaccccaagcgttgaact 600
    |||||||
DB 551 ATCATGATAGGCTCTACAAATTTCAGCAACACTGGTGTACTGACCCCAAGCTGAACACT 610
    |||||||
QY 601 acgtatctcagaagacgtgaagagctgtgcccactgaatggcaacctcagtgcaactgtg 660
    |||||||
DB 611 ACGTATCTCAGACACTGAGAAGGCTGTGTGCCCACTGAATGCAACCTCAGTGCAGTGTG 670
    |||||||
```

```

QY 661 gactltgatctcggagccccaacacatcttcgataaagaactatgtgaactagaggag 720
    |||||||
DB 671 GACTTTGATCTGGCGGACCCCAACCATCTTGATGATTAAGATGATGATGATGATGATGAT 730
    |||||||
QY 721 cagaagggcctgtatagagatgtatcaagaactgtttagcagttccagagcccaactgacac 780
    |||||||
DB 731 CAGAAAGGCTGTATACAGAGTGAATCAAGAACTGTTAGCAGTCCMAAGCAGCAGACACC 790
    |||||||
QY 781 atcccaactgtgaagaatttgcctaactcactcaaacctcttcaagccttcgttgaa 840
    |||||||
DB 791 ATCCCACTGGTGAGAGATTTTGCTAATCTATCACTCAAACTTTTAACTGCTGTGAA 850
    |||||||
QY 841 gccatgg 847
    || ||||
DB 851 GCTATGG 857
```

```

RESULT 3
US-08-378-588-23
; Sequence 23, Application US/08378588
; Patent No. 5608148
; GENERAL INFORMATION:
; APPLICANT: John, Mallyakal E.
; TITLE OF INVENTION: TRANSGENIC COTTON PLANTS
; TITLE OF INVENTION: PRODUCING HETEROLOGOUS PEROXIDASE
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nicholas J. Seay, Quarles & Brady
; STREET: First Wisconsin Plaza, One South
; STREET: Pluckney St.,
; STREET: P.O. Box 2113
; CITY: Madison
; STATE: WI
; COUNTRY: USA
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/378,588
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27,386
; REFERENCE/DOCKET NUMBER: 11-229-9101-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-2484
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1270 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-378-588-23
```

```

Query Match          17.6%; Score 163.4; DB 1; Length 1270;
Best Local Similarity 52.5%; Pred. No. 1.3e-45;
Matches 481; Conservative 0; Mismatches 421; Indels 15; Gaps 5;
```

```
QY 124 ttccatgactgcttcgtgaatggttcgcagcgtacgatalactgagcaaccaccagt 183
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 201 ttccatgattgtttgttaattggttgatgattcaattttgttag---acacagatggc 257
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 184 ttccgacacgaagaagatgattccggaaagcgaacggccgaagggtcttccagatc 243
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 258 actcaaaactgagaaagatgacactgctaa---tgtagctcagagagatttgatattgtg 314
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 244 gatacgatgaaggctgcgcgttcagtcagcatgccacgaacagtcagltg tgcagactg 303
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 315 gatgatattttaaactgcactgacagagaaatgattgacctggtttgattcttgcagaaatt 374
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 304 ctgactatagctgcgcaacagagcgtgactctctgcaggcgagacgttcgtgagagtgccg 363
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 375 tttagcccttgcatctgtaaaatttgagttgtcttgcttgaagagtcgcgtcgagcaagctt 434
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 364 ctgcgtcgaagtgtaacctccctacagagcttccctgactgctgcgaacgcaacttgctc 423
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 435 tttagcagagaagatgactgacacgcaaacccgactgacattatgtagtatttccatgac 494
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 424 ccaattccacccctgcgccagctgaagatagctttagaagctggtgctgaatcgctcg 483
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 495 ccctttgaaaccccttgctgtaattgacgcccacatttcacacaaaggaatgagt---tta 551
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 484 agtgacctgtggtcgtctgctcggaagacacacattggaagaagacagtgtaggttcac 543
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 552 actgattcttgctctgattctgctgacacacatttggaagacaaagattgtagtacttt 611
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 544 atgagataggtctacaaatttcagcaaacactgggttaccctgaccccgctgaacactacg 603
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 612 gaacaaacgttctttaaactcaatgacagtggaattcctgattcaacccgtagacgctaca 671
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 604 tatcccaagacactgagagcgtctgtgcccactgaaatg---caacctgagcactagt 660
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 672 tttttcaaaacattacagaggaatttgcctcaaggtggaatgaatggcaattttttaca 721
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 661 gacttgatctgcggagaccacacactcctcgataacaaagtactatgtgaactcagaggag 720
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 732 aaattgataatatacaactccttaattgatttggatgaatttcactaattcttcaaaat 791
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 721 cagaagaagcctgatacagagtgatacagaacgtttagcagtcagagccactgagcaacc 780
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 792 aatcaaaagactttttcaaaactgattcaagagttgttt---tctacattctgcatctgctaca 848
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 781 atcccaactgtgagaagcttctgtaactctactcaaaccttcttaacgctctgtgaa 840
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 849 attgcatatgataatgatactgctgtagtgcacaaagtcagtttttgatgattttatttgc 908
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 841 gccatgagccgtatgggtataacattaccctctgacgggtacccaaggccaagatcgtctg 900
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 909 tccatgattttaaattgggtatataatgcttaattgctgactaattgagagatttagaaa 968
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 901 aactgcagagtgltcaa 917
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 969 gattgcaagagggttaa 985
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 4  
US-08-811-094-23  
Sequence 23 Application US/08811094  
Patent No. 5869720

GENERAL INFORMATION:  
APPLICANT: John, Maliyakal E.  
TITLE OF INVENTION: TRANSGENIC COTTON PLANTS  
TITLE OF INVENTION: PRODUCING HETEROLOGOUS PEROXIDASE  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nicholas J. Seay, Quarles & Brady  
STREET: First Wisconsin Plaza, One South  
STREET: Pinckney St.,  
CITY: P.O. Box 2113  
CITY: Madison  
STATE: WI

```

      COUNTRY: USA
      ZIP: 53701-2113
      COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      OPERATING SYSTEM: IBM PC compatible
      SOFTWARE: Patent Release #1.0, Version #1.25
      CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/811,094
      FILING DATE: 03-MAR-1997
      CLASSIFICATION: 800
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: 08/378,588
      FILING DATE: 25-JAN-1995
      ATTORNEY/AGENT INFORMATION:
      NAME: Seay, Nicholas J.
      REGISTRATION NUMBER: 27,386
      REFERENCE/DOCKET NUMBER: 11-229-9101-2
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: (608) 251-2484
      TELEFAX: (608) 251-9166
      INFORMATION FOR SEQ. ID NO.: 23:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 1270 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: double
      TOPOLOGY: linear
      MOLECULE TYPE: DNA (genomic)
US-08-811-094-23
```

Query Match 17.6%; Score 163.4; DB 3; Length 1270;  
Best Local Similarity 52.5%; Pred. No. 1.3e-45;  
Matches 481; Conservative 0; Mismatches 421; Indels 15; Gaps 5;

```
QY 4 cagttacccctcaattctacgacaatagctgtcccaagctgtccaaatcgttcgcagc 63
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 81 caattaaaggccaacattttacgattaccacttgccttaattgtaacgaatttgcactggct 140
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 64 acaactgtaacagagctcagatccgataccagatcgcctgctcaattatagctgcagc 123
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 141 gttatgatcaaaagcgaacgactgattgcttcgacgtgctgcttaaaatttgcctttcat 200
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 124 ttccatgactgcttcgtgaatggtltgcagcgtatagatalactggaacaacaccacagt 183
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 201 ttccatgattgtttgtttatggttgatgattcaattttgttag---acacagatggc 257
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 184 ttccgactgaaagaatgcatcctcggaaagcgttaacagcgcgaagggtcttccagtgatc 243
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 258 actcaaaactgagaaagatgacactgcttaa---tgtagtgacagagagattgattttgag 314
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 244 gatacgatgaaggtctgcgttgatgatacgtgcccacgaacgtcgttgtagagcccg 303
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 315 gatgatattttaaactgctactagagaattgattgctggtgtattcttgacgattatt 374
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 304 ctgactatagctgcgcaacagagcgtgactcttgcagcgcggaacgctccgtgagatgccc 363
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 375 tttagcccttgcaattgaaatttgagattgcttgcgttaaaagctccgtgcgcaagacttt 434
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 364 ctgcgtcgaagtgactccctacagagcatctctagatctggtccaaagcgaactgctgct 423
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 435 tttagcagagaagatgactgcttaacagcaaacccgactgagctaaattgattttccttagc 494
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 424 ccaattccacccctgcgccagcgtgaagatagctttagaagaagtggttctgatatgctcg 483
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 495 ccctttgaaaccccttgctgtaattgacgcccaaatccacaaaggaatggatt---tta 551
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 484 agtgacctgtgctcgtcgcggaagacacacattggaagaagaccagtgtaggttcac 543
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 552 actgattcttgcttattctgctgacacacattttggaagacaaagattgtagtactttt 611
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 544 atgagatagctctacaatttcagcaaacactgggttaccctgaccccaagcgtgaacattag 603
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```

Db      612 GAACAACTCTCTTTAACTTCAGTGCAGGTGTAATCTGATTCACAAACCGTAGAGCTTACA 671
OY      604 tatctccagaacacgagagagcttctgtgccactgaatg---caactaagtgcaatg 660
Db      672 TTTTTCACAACTTATACAGGAAATTTGGTCTCAAGGTGGAAATTAAGGCAATATCTTTTACA 731
OY      661 gacttgcctcgcgagaccccaaccatctctcgatacaagaactaagtaagtaagtagag 720
Db      732 AATCTTGATATATCAGCTCCATATGATTTTGTATATATGACTATTTCCATATCTTCAAAAT 791
OY      721 cagaagaagcctgatacagagtgcataagaactglttagcagtcagagcccaatgaacc 780
Db      792 AATCAAGACATCTTCTTCAAACTGATCAACAGTGAAGTGTGT--TCTACATCTGATCTGCTACA 848
OY      781 atccacacggtgagagaagtttgcctaactcactcaaacctctttaaagcccttggaa 840
Db      849 ATTGCTATATAGTAATCGTTATGCTGTGATGTCAAAGTCACTTTTGTGATGATTTTATTTGCG 908
OY      841 gcatatgacacglatatggtaacattaccctctgcagcgggtaccacaagccagattcgtc 900
Db      909 TCGATGATTAATATTGGGTAATATATAGTCCATTAATCTGACTAATATGAGAGATTTAGGAAA 968
OY      901 aactgcagagtgatgca 917
Db      969 GATTGCAAGAGGATTAA 985

RESULT      5
PCT-US94-11121-23
: Sequence 23, Application PC/TUS94111121
: GENERAL INFORMATION:
:   APPLICANT: John, Maliyakal E.
:   TITLE OF INVENTION: TRANSGENIC COTTON PLANTS
:   TITLE OF INVENTION: PRODUCING HETEROLOGOUS PEROXIDASE
:   NUMBER OF SEQUENCES: 23
:   CORRESPONDENCE ADDRESS:
:     ADDRESSEE: Nicholas J. Seay, Charles & Brady
:     STREET: First Wisconsin Plaza, One South
:     STREET: Pluckney St.,
:     STREET: P.O. Box 2113
:     CITY: Madison
:     STATE: WI
:     COUNTRY: USA
:     ZIP: 53701-2113
:   COMPUTER READABLE FORM:
:     MEDIUM TYPE: Floppy disk
:     COMPUTER: IBM PC compatible
:     OPERATING SYSTEM: PC-DOS/MS-DOS
:     SOFTWARE: Patentin Release #1.0, Version #1.25
:   CURRENT APPLICATION DATA:
:     APPLICATION NUMBER: PCT/US94/11121
:     FILING DATE:
:   CLASSIFICATION:
:     ATTORNEY/AGENT INFORMATION:
:       NAME: Seay, Nicholas J.
:       REGISTRATION NUMBER: 27,386
:       REFERENCE/DOCKET NUMBER: 11-229-9076-8
:     TELECOMMUNICATION INFORMATION:
:       TELEPHONE: (608) 251-2484
:       TELEFAX: (608) 251-9166
:     INFORMATION FOR SEQ ID NO: 23:
:       SEQUENCE CHARACTERISTICS:
:         LENGTH: 1270 base pairs
:         TYPE: nucleic acid
:         STRANDEDNESS: double
:         TOPOLOGY: linear
:     MOLECULE TYPE: DNA (genomic)
PCT-US94-11121-23

```

OY	4	caatgaacccctacatctcctaaagaacatagctgctcccaagctgccaacatcgtctcgac	63
Db	81	CAATTAAAGCGCAACATTTTACGATATCCACATCTGCCCTTAATGTTACAAAGTATTGTACGTGGT	140
OY	64	acaatcgtcaacagagctcagaatcccatcccaaggaatcgctgcttcctcaatatacgtctcgac	123
Db	141	GTTATGATCAAAAGGCAACGACTACGATGCTGCAGCTGGTCTAAATTAATTAATTCGCTTCAT	200
OY	124	ttccatgaactgcttcggaataggtttgcagcgctcagatatactactggaacaacaccacagat	183
Db	201	TTCCATGATATTTGTTTAATAGGTTGGATGATCAATTTTGTTAG--ACACAGATGGG	257
OY	184	ttccgcacatgaanaagatgacatccgggaacgctaacagcgccaggggctctccagatgac	243
Db	258	ACTCAAACTGGAAAAGATGACACCTGCTAA---TGTAGTGCAGAGAGATTTGATTTTGTG	314
OY	244	gacgcacatgaagagctgcgcgtttagtcagcatgcccacgaaacagttcagtttgcagacgt	303
Db	315	GATGATATTAAACAGTCGACCTAGAAATGATGATCCCTGTTGATCTGTGCAGATATT	374
OY	304	ctgaactatagctgcgaacagagcgttactcttgcagcggaacgcttctgtagaagtcg	363
Db	375	TTAGCCCTTGGATCTGAAATTTGGAGTTGTCTTGGCTAAAGCTCCGTCGTGGCAAGTACTT	434
OY	364	ctcgcgtcagctgacatccctcaacagacatccctagatcttgcgaacgccaacgttcgct	423
Db	435	TTTGGCAGAAAGATAGTATTACAGCAACACGATCTGGACTTAATAGTATTCTCAGC	494
OY	424	ccatctctcacccctgcgcagcgtgaagatagctttagaacgtybggtctgaatcgctcg	483
Db	495	CCCTTTTGAACCCCTTGGCTGTATAGCCACACAAATTTCCACCAAGGGAATGAT--TTA	551
OY	484	agtagacctgtgctgtcgtccggaggacaacatttggaaagaacccagttagttcatc	543
Db	552	ACTGATCTTGTGCTCTATCTGCTGACACACATTTTGAAGAGCAGATGTGGTACTTTT	611
OY	544	atggaatgaagctcaaatctcagaacaacatggttactctgaaccaccaagctgaaacaatag	603
Db	612	GAACACGCTCTCTTAACTTCAAGTGGCAGTGTAACTGATCCAACGCTAAGACGTACA	671
OY	604	tatctcagaacacatgagagcttctgcccactgaatgg--caacctagtgcaatagtg	660
Db	672	TTTTTACCAACATTTACAAGGAATTTGTCTCAAGGTGGAAATATAGCAATACTTTTACA	731
OY	661	gaacttgtctgcggagaccaccaatctctgatacaagaactactatgaaatccagaggag	720
Db	732	AATCTTGATATATCAACCTCCTAATGATTTTGTATATGATATATTTCACTAATCTTCAAAAT	791
OY	721	cagaaagcccgatatacagagtgatcaagaagcttttagcagtcgacagccacatgacac	780
Db	792	AATCAAGACCTCTTTCAAACATGATCAAGAGATGTTT---TCTACATCTGGATCTGCTACA	848
OY	781	atcccaactggtgagaaagtttctgtaactactactcaaaacctcttlaagcgcctctgtgaa	840
Db	849	ATTGCTATATAGCAATTCGTTATATGCTGTAGTCAAGTCAAGTTGATTTTGAATTTATTTGC	908
OY	841	ggcaatgagccglatagggtaaacattaacccctctgcagcggtatcccaagccagatctgtc	900
Db	909	TTCGATGATTTAAATTTGGTAAATATATAGTCCATTAATCTGTACTAATGAGAGATTAGAAA	968
OY	901	aactgcagagctgtcaaa	917
Db	969	GATTGCAGAGAGGTTAA	985
RESULT			
US-08-190-029A-11			
: Sequence 11, Application US/08190029A			
: Patent No. 5736363			
: GENERAL INFORMATION:			
: APPLICANT: EDWARDS, Richard Mark			
: APPLICANT: BAWDEN, Lindsey			
:			

```

1  TITLE OF INVENTION:  IGF-II ANALOGUES
2  NUMBER OF SEQUENCES:  12
3  CORRESPONDENCE ADDRESS:
4  ADDRESSEE:  ALLEGRETTI & WITCOFF, LTD.
5  STREET:  10 S. WACKER DRIVE, SUITE 3000
6  CITY:  CHICAGO
7  STATE:  ILLINOIS
8  COUNTRY:  U.S.A.
9  ZIP:  60606
10
11  COMPUTER READABLE FORM:
12  MEDIUM TYPE:  Floppy disk
13  COMPUTER:  IBM PC compatible
14  OPERATING SYSTEM:  PC-DOS/MS-DOS
15  SOFTWARE:  Patentin Release #1.0, Version #1.25
16  CURRENT APPLICATION DATA:
17  APPLICATION NUMBER:  US/08/190.029A
18  FILING DATE:  28-FEB-1994
19
20  CLASSIFICATION:  435
21  PRIOR APPLICATION DATA:
22  APPLICATION NUMBER:  PCT/GH92/01389
23  FILING DATE:  27-JUL-1992
24  PRIOR APPLICATION DATA:
25  APPLICATION NUMBER:  GB 9202401.7
26  FILING DATE:  05-FEB-1992
27
28  PRIOR APPLICATION DATA:
29  APPLICATION NUMBER:  GB 9116325.3
30  FILING DATE:  29-JUL-1991
31  ATTORNEY/AGENT INFORMATION:
32  NAME:  JOHN J. McDONNELL
33  REGISTRATION NUMBER:  26,949
34  REFERENCE/DOCKET NUMBER:  94,062
35  TELECOMMUNICATION INFORMATION:
36  TELEPHONE:  312-715-1000
37  TELEFAX:  312-715-1234
38  INFORMATION FOR SEQ ID NO:  11:
39
40  SEQUENCE CHARACTERISTICS:
41  LENGTH:  388 base pairs
42  TYPE:  nucleic acid
43  STRANDEDNESS:  double
44  TOPOLOGY:  linear
45  MOLECULE TYPE:  DNA (genomic)
46  ORIGINAL SOURCE:
47  ORGANISM:  Homo sapiens
48
49  FEATURE:
50  NAME/KEY:  CDS
51  LOCATION:  11..373
52  OTHER INFORMATION:  /codon_start= 11
53  OTHER INFORMATION:  /function= "Gene for IGF-II/HRP fusion protein"
54  OTHER INFORMATION:  /product= "Synthetic DNA"
55
56  FEATURE:
57  NAME/KEY:  misc_feature
58  LOCATION:  1..6
59  OTHER INFORMATION:  /function= "HindIII cleavage site"
60
61  FEATURE:
62  NAME/KEY:  misc_feature
63  LOCATION:  383..388
64  OTHER INFORMATION:  /function= "EcoRI cleavage site"
65
66  FEATURE:
67  NAME/KEY:  misc_feature
68  LOCATION:  11..172
69  OTHER INFORMATION:  /function= "HRP coding sequence"
70
71  FEATURE:
72  NAME/KEY:  misc_feature
73  LOCATION:  173..373
74  OTHER INFORMATION:  /function= "IGF-II coding sequence"
75
76  OS-08-190-029A-11

```

Query Match	17.2%;	Score 159.4;	DB 2;	Length 388;
Best Local Similarity	99.4%;	Pred. No. 1.4e-44;		
Matches 160;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;

```

Db      11  ATGCAGTTAACCCTACTATTCTAGACAAATAGCTGTCCCAAGCTGCACATTCGTTCCG 70
          |||||||
Oy      61  gacacaaatcgctcaacagagctcagatccgataccgatacgcgtctgtcctaataattacgctg 120
          |||||||
Db      71  GACACAAATCGTCAACAGAGACTCAGATCCGATCCGATCCCGATCCGTCGTTCAATATTACGCTG 130
          |||||||
Oy      121 cacttccatgactgctcttcgtgaaatggttgcgagctgcat 161
          |||||||
Db      131 CACTTCATWAGACTGCTTCGTGTAATGGTTGCACGCTAGTAT 171
          |||||||

RESULT 7
US-08-462-695-11
; Sequence 11, Application US/08462695
; Patent No. 5854025
; GENERAL INFORMATION:
; APPLICANT: EDWARDS, Richard Mark
; APPLICANT: BARDEN, Lindsey
; TITLE OF INVENTION: IGF-II ANALOGUES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BANNER & ALLEGRETTI, LTD.
; STREET: 10 S. WACKER DRIVE, SUITE 3000
; CITY: CHICAGO
; STATE: ILLINOIS
; COUNTRY: U.S.A.
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,695
; FILING DATE: 5-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/190,029
; FILING DATE: 28-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB92/01389
; FILING DATE: 27-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9202401.7
; FILING DATE: 05-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9116325.3
; FILING DATE: 29-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: JOHN J. MCDONNELL
; REGISTRATION NUMBER: 26,949
; REFERENCE/DOCKET NUMBER: 94,062-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-715-1000
; TELEFAX: 312-715-1234
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 388 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 11..373
; OTHER INFORMATION: /codon_start= 11
; OTHER INFORMATION: /function="Gene for IGF-II/HRP fusion protein"
; OTHER INFORMATION: /product="Synthetic DNA"
; NAME/KEY: misc_feature

```

LOCATION: 1..6  
OTHER INFORMATION: /function- "HindIII cleavage site"  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 383..388  
OTHER INFORMATION: /function- "EcoRI cleavage site"  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 11..172  
OTHER INFORMATION: /function- "HRP coding sequence"  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 173..373  
OTHER INFORMATION: /function- "IGF-II coding sequence"  
US-08-462-695-11

Query Match 17.28; Score 159.4; DB 3; Length 388;  
Best Local Similarity 99.4%; Pred. No. 1.4e-44;  
Matches 160; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atgcagtaacccttaattctagacaatagctgtcccaacgtgtccaatctgttcgc 60  
|||||  
Db 11 ATGAGTTAACCCCTACATTCTACGACAAATAGCTGCCCAACGTGCCAATCGTTCGC 70  
61 gacacaatcgtcaacgagccagatccgaltcccgatcgtcttcaatatacgtctg 120  
|||||  
Db 71 GACACATCTCTCAACGAGCTCAGATCCGATCCGATCCGATCCGTTCAATATTCGCTG 130  
121 cacttcacatgactgctcgtgaatggtgacgctagcat 161  
|||||  
Db 131 CACTTCATGATGCTCTCGTGAATGTTGGAGCGCTGATAT 171

RESULT 8  
US-08-181-271A-22  
Sequence 22, Application US/08181271A  
Patent No. 5614395  
GENERAL INFORMATION:

APPLICANT: Ryals, John A.  
APPLICANT: Alexander, Danny C.  
APPLICANT: Beck, James J.  
APPLICANT: Duesing, John H.  
APPLICANT: Friedlich, Leslie B.  
APPLICANT: Goodman, Robert M.  
APPLICANT: Harms, Christian  
APPLICANT: Meins, Jr., Frederick  
APPLICANT: Montoya, Alice  
APPLICANT: Moyer, Mary B.  
APPLICANT: Neuhaus, Jean-Marc  
APPLICANT: Payne, George B.  
APPLICANT: Sperison, Christoph  
APPLICANT: Stinson, Jeffrey R.  
APPLICANT: Uknes, Scott J.  
APPLICANT: Ward, Eric R.  
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: New York  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/181,271A

FILING DATE: 13-JAN-94  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/093,301  
FILING DATE: 16-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/937,197  
FILING DATE: 6-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/678,378  
FILING DATE: 1-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/305,566  
FILING DATE: 6-FEB-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/165,667  
FILING DATE: 8-MAR-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/042,847  
FILING DATE: 6-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/632,441  
FILING DATE: 21-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/425,504  
FILING DATE: 20-OCT-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/848,506  
FILING DATE: 6-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/580,431  
FILING DATE: 7-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/368,672  
FILING DATE: 20-JUN-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/329,018  
FILING DATE: 24-MAR-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/045,957  
FILING DATE: 12-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8614  
TELEFAX: (919)541-8689  
INFORMATION FOR SEO ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1131 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-181-271A-22

Query Match 16.0%; Score 148.6; DB 1; Length 1131;  
Best Local Similarity 50.7%; Pred. No. 1.3e-40;  
Matches 464; Conservative 2; Mismatches 431; Indels 18; Gaps 4;

QY 4 cagltaaccccttaattctagacaatagctgtcccaacgtgtccaatctgttcgc 63  
|||||  
Db 98 CAGTTTCTCTCTTCTCTTACGACCAATAGCTGCCAATCGTTCGTTCTCAAC 157  
64 acaatcgtcaacgagctcagatccgaltcccgatcgtcttcaatatacgtctgac 123  
|||||  
Db 158 GTGTGTCGCCAAGCCCTACAGACTGATGACCGAGACTGCTCTAGCTCATTCGCCCTCAT 217





TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8614  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1131 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-449-315-22

Query Match 16.0%; Score 148.6; DB 1; Length 1131;  
Best Local Similarity 50.7%; Pred. No. 1,3e-40;  
Matches 464; Conservative 2; Mismatches 431; Indels 18; Gaps 4;

```
OY 4 caglttaacccatcatctacgacaatagctgtcccaacgtgtcccaacatcgttcgcgac 63
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 98 CAGYTTCTCTCTTCTTCTTACGCCACACATGCCCTCAGCTGCTTGGTTGTTCTCAC 157
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 64 acaatgtaacgagcctcagatcccgatccagatcgtcgttcaataatcgttcgac 123
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 158 GTGGTCCCAAGCCCTACAGACTGATGACCGAGCTGCTGTAAGCTATTCGCTCCAT 217
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 124 ttccatgactcgttcgtgagctgagcgtacgattactggaacaacaccagc 183
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 218 TTTTCATATTTGCTTTTGTCAATGGGTGATGATGATGATTCATTGTTGACGTACGGGC 277
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 184 ttccgcactgaaagatgcatcgggaacgttaacgcgcgaagggtcttcacgtatc 243
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 278 GTATATCATATGTAATTAATGACCTCCAAATGTTGGAATCCAAAGAAATGACATTGTG 337
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 244 gatcgatgagcgtcgttcgtgagctgagcgtacgacgaaacgtcgttcgacgactg 303
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 338 GACAACATCAAGCAGAGATTGAGAGTGTTCACAGAGTGTTCCTGCGCTGATATC 397
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 304 ctgactaagctgcgcacagagcgtgacctgtgcagcgagaccgtctcgtgagagtcgcg 363
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 398 TTAGCCATTTCATCTCAATCTCTGTTTCTGTCGGGAGGACCAATTTGGTTGTACCA 457
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 364 ctgcgtgacgtgacctccctcaccagcatctctatagctggtcccaacgttcgtcgtc 423
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 458 ATGGGAAGAAAGACAGCAGATATGCAATAGAACT---GGAACCTCAAACTTACCTGGGT 514
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 424 ccatcttcacacctcccccagcgtgaaagatagctttagaaacgttggtctgtaacgtcgcg 483
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 515 CCTCAGAAACTCTAGTGGGACTTAAAGGCATTTAAAGATCAAGGGCTTGAT---TCT 571
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 484 agtgaacttggtgctgtcgcgagagacacacatttggaagaacacagtgtagtctatc 543
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 572 ACAGATCTCGGGCTCTATACAGGACCCACACGTTTGAATAATCAAGATCATCTTCTTC 631
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 544 atgatatgctctacaattcaagaacacgtggttacctgaccccaacgtcgaacactacg 603
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 632 AGTGAAGCGCTCATCACTTCAAGGACAGGAAGACCCGACACAAAGCGTTGACCCATA 691
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 604 tatctcagaacactgagagcgtgtgtcccaactgaatggaacacctcagtgacactagtgac 663
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 692 TACAGGAGACAGCTTTCAGAGCTTTGTACTACTCAACAAA-----CACGAGTAAT 742
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 664 ttgatctgagagacccaacatctctgataaagaagtaactgtgaatctagagaagacag 723
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 743 TTTCACCCAGCTACACCCACTAGATTGACAGAGCTTATTACAACTTGAATTTAGTTAGCTTA 802
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 724 aaagcctgtatagatgtaacaaactgtttagcagtcgaagcgcacactgacacatc 783
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 803 AGAGGGCTTCTCCAAAGCAGCAAGAGCTCTTCTCAACTCC---CAGAGCTGATACACCA 859
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 784 ccactgtgtgagaagtttctgtaactactcaaacctctcttaagcgttcgtgtgaaagcc 843
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 860 GCCATTGTCAAACTTTGCTGCGCAAGCAAGCTGCTTCTTAAACAATTTGTGAATCA 919
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
OY 844 atggaccgtatggtaacattaccctctgacgggtacccaagccagatctgtcgtgac 903
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 920 ATGATCAAAATGGGCAACCTCAAGCCCTCCCTGCGCATTCGATTCAGAACTTACATTGAC 979
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 904 tgcagagtggtcaac 918
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 980 TGTAAAGAGGTCAAC 994
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

## RESULT 10

US-08-444-803-22

Sequence 22, Application US/08444803

Patent No. 5654414

GENERAL INFORMATION:

APPLICANT: Ryals, John A.

APPLICANT: Alexander, Danny C.

APPLICANT: Beck, James J.

APPLICANT: Duesling, John H.

APPLICANT: Friederich, Leslie B.

APPLICANT: Goodman, Robert M.

APPLICANT: Harms, Christian

APPLICANT: Meins, Jr., Frederick

APPLICANT: Montoya, Alice

APPLICANT: Moyer, Mary B.

APPLICANT: Neuhaus, Jean-Marc

APPLICANT: Payne, George B.

APPLICANT: Sperison, Christoph

APPLICANT: Stinson, Jeffrey R.

APPLICANT: Uknes, Scott J.

APPLICANT: Ward, Eric R.

APPLICANT: Williams, Sherilca C.

TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC

NUMBER OF INVENTION: DNA SEQUENCES AND USES THEREOF

CORRESPONDENCE ADDRESS: 106

ADDRESS: CIBA-GEIGY Corporation

STREET: 7 Skyline Drive

CITY: Hawthorne

STATE: New York

COUNTRY: USA

ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/444,803

FILING DATE: 19-MAY-1995

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/181,271

FILING DATE: 13-JAN-94

APPLICATION NUMBER: US 08/093,301

FILING DATE: 16-JUL-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/937,197

FILING DATE: 6-NOV-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/678,378

FILING DATE: 1-APR-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/305,566

FILING DATE: 6-FEB-1989

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/165,667

FILING DATE: 8-MAR-1988

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/042,847

FILING DATE: 6-APR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/632,441

FILING DATE: 21-DEC-1990

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/PI/CCG 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8614
INFORMATION FOR SEO ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 1131 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-444-803-22
```

```
Query Match          16.0%; Score 148.6; DB 1; Length 1131;
Best Local Similarity 50.7%; Pred. No. 1.3e-40;
Matches 464; Conservative 2; Mismatches 431; Indels 18; Gaps 4;
```

```

QY 4 cagttaacccatctctagacagacatgctgtcccaagctgtcccaacatcgcttcgcgac 63
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 98 CAGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 157
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 64 acaatgcgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 123
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 158 GTGGTTCGCCAGCCCTACAGACTGATGACCGAGCTGCTCTAAGCTTCATTCGCTTCAT 217
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 124 ttccatgactgcttcgtgaatgttgacagcctagacatctactgagacacaccacagc 183
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 218 TTTCATGATGCTTGTGTCATGGGTGATGATGATGATGATGATGATGATGATGATGATG 277
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 184 ttccgaactgaaagatgcatcctgcggagcgtacagcgccgagcgccgttcagtgatc 243
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 278 GTATCATGATGTAATCTTATGAGACTCCAAATGCTGGAATCCAAAGAAATGACATGTG 337
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 244 gacgcagaagcgtgcgcttgatcagcagctgcccacgacacagcagcagcagcagcagc 303
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 338 GACAACATCAACAGCAGCAGTGTGAGAGCTTGTCACAGCAATGTTCTTCTCGCTGATATC 397
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 304 ctgactatactgctgcgaacagacgctgactctgcagcgagcagcgtctcctgagagatgcg 363
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 398 TTAGCATTTTCATCTCAAAATCTCTGTTTCTGTGTGCGGAGAGACCAATTTGGGTGTACA 457
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 364 ctgcgtgcagcgtgactccctacagcgaatccttgatcctgacatgacgacacagcagc 423
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 458 ATGGGAAGAAAGACAGCAGCAATATGACCAATATGAACT--GGAACCTCAAACTTACCTG 514
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 424 ccaattcaccctgcgcacgctgaagatagctttagaagcgtgagctgagctgagcagc 483
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 515 CCCTCAGAAACTCTAGTGGAGCTTAAAGCAAGTTTAAAGATCAAGGCGTTGAT--TCT 571
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```

QY 484 agtaccctgtggtcctctgtccgagagacacacatlttgaaagaaccagtgtaggtcacc 543
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 572 ACAGATCTGCTGGCTCTATCAGAGAGCCACACGTTTGGAAATCAAGATGCATGTTCTTC 631
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 544 atggatagccttcaaatctcagcaaacacttgggttaccttgcacccagcgtgaacactag 603
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 632 AGTGACCGCTCATCAACTTCAACGGCAGAGAGACCCGACACACGCTTGAACCCATA 691
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 604 taticcagacacactgaaagagcctgtgtgcacactgaatgacacctcagtgacactagtgac 663
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 692 TACAGGAGCAGCTTGCAGACATTTGACTACTCAACAAA-----CACGATTAAT 742
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 664 ttgatctgcgagcccccacacatcttgataaacagactatgtgaatctagaagagcag 723
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 743 TTGCACCCAGTCAACCCACTAGATTGACAGACCTATTACAAATTTGATTAGTTA 802
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 724 aaagcctgataagatgtgtcaagaactgtttaagcagctcagacgcaactgacacatc 783
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 803 AGAGGGCTTCTCCAAAGCGACCAAGACCTCTTCAACTCC---CAGAGCTGATACCACA 859
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 784 ccaactgtgagaagcttctgctactctactcaaacctctttaaagcctcgtgcgagacc 843
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 860 GCATTGTCAAACTTTTGTGTCGCCACAGACGTCCTCTTTTAAACATTTGTGAATCA 919
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 844 atgagccgtatggtaacattaccctctgacggtacccaagcgacatctcgtcgtgaac 903
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 920 ATGATCAAAATGGGCAACCTCAAGCCCTCCCGGCTTGTCAATCAGAAATTAATTGAGAC 979
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 904 tgcagagtggtcaac 918
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 980 TGTAAAGCGCTCAAC 994
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```

RESULT 11
US-08-449-043-22
Sequence 22, Application US/08449043
Patent No. 5689044
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesling, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Umes, Scott J.
APPLICANT: Ward, Eric R.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESSES:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/449,043
```



```

STREET: 520 White Plains Road, P.O. Box 2005
City: Tarrytown
STATE: New York
COUNTRY: USA
ZIP: 10591

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/456,265A
FILING DATE: 31-MAY-95
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/181,271
FILING DATE: 13-JAN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT 1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727/DIV1010
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8587
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 1131 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA

```

US-08-456-265A-22

Query Match	16.0%;	Score 148.6;	DB 2;	Length 1131;
Best Local Similarity	50.7%;	Pred. No. 1.3e-40;		
Matches 464; Conservative	2;	Mismatches 431;	Indels 18;	Gaps 4;

QY	4	caagttacccctatattcttcaagaaatagtctgtcccaagtggtccaaatctgtctgcgaac	6
Db	98	CAGTATTTCCTCTTCTCTCTAAGCCACACATGCCCTCAGCTGCCCTTTCGTGTTCACAC	157
QY	64	acaalgcgtcaacgagctcagatccgaatcccaaggatgcgtcttcataatcagtcgtcac	123
Db	158	GTGTTGTGCCCAAGGCCCTACAGACTGATGAGCCGAGCTCTCTAAGCTCATTTCCCTCCAT	217
QY	124	ttcatgtactgtcttcgtgaatagttgtgagcgtcagatattactgtgaaacaccaccagt	183
Db	218	TTTTCATGATTTCCTTTGTGCATAGGGGTGTGATGATCGATTCTATTGTGGAGCGTACCGGGC	277
QY	184	ttccgcactgaaaaagatgcatcttcgggaaagcttaacagcgccagggggcttccagtgatc	243
Db	278	GTTTATTCGATTACTGAACCTTAATGACGCTCCAAATGCTGGAAATCCAGGAATGACATTTGTG	337
QY	244	galtcgcataagagctgtccggtttagtcagatgcccacgaaacagttgtagtcagacctg	303
Db	338	GACACATCTCAAAAGCAGAGATTGGAAGTGTGTCTCCAGAGATGTCTTCTTCGCTGTGATTC	397
QY	304	ctgacataagctgtgcacaacgagagtgtagctcttgccagcgagaccgtctctgagagtgccg	363
Db	398	TTTAGCCATTTCATCTCAAAATCTCGTTTCTGTGTGGAGGAGCCAAATTTGGTTGTACCA	457
QY	364	ctcgtctgcagctgtagctccctccatacaggaatccctagatctggtccaaagccaaactgtgcct	423
Db	458	ATGGGAAGAAAAGACACACAGATATGCCAAATAGAAGCT---GGAACCTCAAACTTACCTCGGT	514
QY	424	ccattcttaccgcccgcccccgactggaagatagatcttagaagacgtgtgtgtcgaatgctcg	483
Db	515	CCCTTCAGAAATCTTAGTGGGACTTAAAGGCAAGTTTAAATATCAAGGGCTGTGAT---TCT	571
QY	484	agtgtacctgtgtgtctgtctcgcgagagacacacatttgaagaagacagtgtagttctatc	543
Db	572	ACAGATCTCGTGCGCTTATCATCAGGAGGCCACACGTTTGGAATATCAAGATGCATGTTCTTC	631
QY	544	atgtaataggtctctcaatttccagcaaacactcgtggtttacctgtgaaccccaagctgaaactacg	603
Db	632	AGTGACCGCCTCATCACTTTCACAGGACACAGGAAGACCCACACAAAGCTGTGACCCCAATA	691
QY	604	tatctccagaacactgaaagagctgtgtgccactgaatgacaactcagtgcaactgtggagc	663
Db	692	TACAGGAGAGCGCTTGCAAGACTTTGTACTACTTCACAAAT-----CACAGATTAAT	742
QY	664	ttgtatctgtcgagaccccaacatcttcgtataacaagtaactgtgaaactctagaagagcag	723
Db	743	TTTCACCCAGTCAACCCCACTTAGATTGTGACAAACATTCATCAAAATTTGATTACTCTTA	802
QY	724	aaagagcctgtatcacagagtgatcaaagacgtttagaagttccaaagcccaactgtgacacatc	783
Db	803	AGAGGGCTTCTCCAAACGACCAAGAGACTTCTCTCAACTCC---CAGAGCTGATTCACCA	859
QY	784	ccactgtgtgaagaagtttgttaactcttaactcaaaccttcttlaagcgtcttgcggaagcc	843
Db	860	GCCATTGTCAAAACTTTTGTGTGCCAAGCAAGCGTGGCTTTTAAACAAATTTGTGAATCA	919
QY	844	atggaacgtaatggttaacatcattaccctctgaagcggttaaccaagagcagatctgctgaac	903
Db	920	ATGATTCAAAATTTGGCAACCTCAAGAGCTCCCTCCGTGGCATTTGCATCAAGAAATTGATTTGAC	979
QY	904	tgcagagtggtcaac	918
Db	980	TGTAGAGGGTCAAC	994



```
Db 692 TACAGGAGCAGCTTCGACAGCTTTGACTACTACACAAA-----CAGGAGTAAT 742
QY 664 ttgatctggagccccaaccatcttgataaagaagactatgtgaatcagagggagcag 723
Db 743 TTGAGCCCACTACACCCACTAGATTGACACAGACCTATTACAACAATTTGATTAGCTTA 802
QY 724 aaagccctgaatacagagtgaatcaagaactgttagcagtcacagcgcacatgcacacatc 783
Db 803 AGAGGGCTTCTCCAAAGCGACCAAGAGCTCTTCTGCAACTCC---CAGAGCTGATACACACA 859
QY 784 ccaactgtgagaagtttctgaactactcaaacctcttlaagccttcgtggaagcc 843
Db 860 GCCATTGTCTCAAACTTTGCTGCCAAGCAAGCAAGCTCTTTTAAACAATTTGTGCAAAATCA 919
QY 844 atgaccctgttggtgaatacattacccctcgtgaggtacccaagccagatctctgaac 903
Db 920 ATGATCAAAATGGGCAACCTCAAGCCTCCCTTGCGCAFTTGCAATCAGAAATTAGTTGGAC 979
QY 904 tgcagagtgtgaac 918
Db 980 TGTAAAGAGGCTCAAC 994
```

```
RESULT 14
US-08-455-244-22
Sequence 22, Application US/08455244
Patent No. 5789214
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericea C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/455,244
FILING DATE: 31-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
```

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/Pl/CCG 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 1131 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-455-244-22
```

```
Query Match 16.0%; Score 148.6; DB 2; Length 1131;
Best Local Similarity 50.7%; Pred. No. 1.3e-40;
Matches 464; Conservative 2; Mismatches 431; Indels 18; Gaps 4;

QY 4 caattacccttaccattcttaccgaatagctgtcccaagctgtccaaatcgaatcgttcgagc 63
Db 98 CAGTTCCTTCCTTCTTCTTACGCCACACATGCTCCAGCTGAGTGCCTTTCGTTGTTCTCAAC 157
QY 64 acaatcgtlaacagagctcagatccgatcccaagatcgtctgtcctaatalatcagatcgcac 123
Db 158 GTGGTTGCCCAAGCCCTACAGACTGATGACCGAGCGTGTCTTAAGCTCATTCGCCCTCCAT 217
QY 124 ttccatgaactcgtctgtaatgtgtcagcagctagacatattactgacacaacaccacagt 183
Db 218 TTTCATGATTCCTTGTTCATATGGGTGTGATGATTCATTTATTGGTAGACGTAACCGGCGC 277
QY 184 ttccgacgtgaaagatgtcattccggaacgcctaacaagcgccaggggcttccagtgatc 243
Db 278 GTTATGATGATGACTTAATGAGCTCCAAATGTGTGAATCCAAAGAAATGACATTTGG 337
```







---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2000, 05:46:18 ; Search time 1168.17 Seconds  
(without alignments)  
3499.414 Million cell updates/sec

Title: US-09-246-451-16

Perfect score: 927

Sequence: 1 atgcagtaacctacatc.....gagtgtaacagcaactct 927

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues

Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*

44: gb\_est25:\*  
45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: em\_est20:\*  
50: em\_est21:\*  
51: em\_est22:\*  
52: gb\_est30:\*  
53: gb\_est31:\*  
54: gb\_est32:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: em\_est35:\*  
83: em\_est36:\*  
84: em\_est37:\*  
85: gb\_est48:\*  
86: gb\_est49:\*  
87: gb\_est50:\*  
88: gb\_est51:\*  
89: gb\_est52:\*  
90: gb\_est53:\*  
91: gb\_est54:\*  
92: gb\_est55:\*  
93: gb\_gss1:\*  
94: gb\_gss2:\*  
95: gb\_gss3:\*  
96: gb\_gss4:\*  
97: em\_gss1:\*  
98: em\_gss2:\*  
99: em\_gss3:\*  
100: em\_gss4:\*  
101: gb\_gss5:\*  
102: gb\_gss6:\*  
103: em\_gss5:\*  
104: em\_gss6:\*  
105: gb\_gss7:\*  
106: gb\_gss8:\*  
107: gb\_gss9:\*  
108: em\_gss7:\*  
109: em\_gss8:\*  
110: gb\_gss11:\*  
111: gb\_gss10:\*  
112: em\_gss9:\*  
113: em\_gss10:\*  
114: em\_gss11:\*  
115: em\_gss12:\*  
116: gb\_gss12:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C	1	290	31.3	620	62	AV441260
	2	276.6	29.8	516	62	AV442222
	3	236.4	25.5	720	122	CNSOOWXO
	4	234.8	25.3	652	118	AQ956990
	5	210	22.7	564	20	AA043293
	6	204.6	22.1	537	29	AA650756
	7	204.4	22.0	824	46	AI1959837
	8	187	20.2	499	90	T88184
	9	185.4	20.0	454	62	AA416100
	10	184.2	19.9	457	87	N37495
C	11	176.2	19.0	461	85	H37375
	12	172.4	18.6	625	72	AM424453
	13	172.2	18.6	691	70	AM278775
	14	169.6	18.3	745	74	AA5595660
	15	164.4	17.7	360	89	T039699
	16	161.6	17.4	387	62	AV439974
	17	159.8	17.2	688	74	AA585492
	18	158.4	17.1	497	25	AA394440
	19	154.4	16.7	308	89	T45859
	20	153	16.6	485	89	T42100
C	21	153.8	16.6	668	74	AA584286
	22	151.6	16.6	638	74	AA587236
	23	151.6	16.4	672	69	AA1857369
	24	151.2	16.3	321	85	F15487
	25	150	16.2	592	80	AM704659
	26	149.8	16.2	640	74	AA584357
	27	149.2	16.1	595	92	T29133
	28	148	16.0	602	80	AA705617
	29	147.8	15.9	479	89	T42049
	30	147.8	15.9	658	79	AA685235
C	31	147.6	15.9	584	80	AM705730
	32	146.4	15.8	607	44	AI182928
	33	146.2	15.8	632	74	AA584356
	34	145.2	15.7	639	70	AA556487
	35	145.2	15.7	646	70	AA267813
	36	144	15.5	475	89	T22196
	37	144	15.5	527	118	AQ955253
	38	142.8	15.4	659	74	AA559945
	39	139.4	15.0	608	80	AM705946
	40	138.4	14.9	668	79	AA685437
C	41	136.6	14.7	583	46	AI1960823
	42	136.4	14.7	670	74	AA574244
	43	136.4	14.7	690	74	AA586142
	44	135.8	14.6	377	30	AA728602
	45	135.4	14.6	703	74	AA576032
C	46	135.4	14.6	703	74	AA576032
	47	135.4	14.6	703	74	AA576032
	48	135.4	14.6	703	74	AA576032
	49	135.4	14.6	703	74	AA576032
	50	135.4	14.6	703	74	AA576032
	51	135.4	14.6	703	74	AA576032
	52	135.4	14.6	703	74	AA576032
	53	135.4	14.6	703	74	AA576032
	54	135.4	14.6	703	74	AA576032
	55	135.4	14.6	703	74	AA576032

## ALIGNMENTS

RESULT	1
AV441260/c	
LOCUS	
AV441260	620 bp mRNA
AV441260	EST
old Arabidopsis thaliana	19-APR-2000
Arabidopsis thaliana above-ground organ two to six-week	
old Arabidopsis thaliana cDNA clone AP237e04_f 3', mRNA sequence	

ACCESSION	AV441260
VERSION	AV441260.1
KEYWORDS	GI:7611644
SOURCE	EST.
ORGANISM	Thale cress.
REFERENCE	Arabidopsis thaliana
AUTHORS	Eukaryota: Viridiplantae; Embryophyta: Tracheophyta; Spermatophyta: Magnoliophyta: eudicotyledons; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
TITLE	1 (bases 1 to 620)
JOURNAL	Nakamura,Y. and Asamizu,E.
COMMENT	Arabidopsis thaliana expressed sequence tags Unpublished (2000) On Jul 7, 1999 this sequence version replaced gi:5410020. Contact: Erika Asamizu The First Laboratory for Plant Gene Research Kazusa DNA Research Institute Yama 1532-3, Kisarazu, Chiba 292-0812, Japan Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/. Location/Qualifiers
FEATURES	1..620
source	/organism="Arabidopsis thaliana" /strain="Columbia" /db_xref="taxon:3702" /clone="AP23/e04.f" /clone_1id="Arabidopsis thaliana above-ground organ two to six-week old" /tissue_type="above-ground organ" /dev_stage="two to six-week old"
BASE COUNT	178 a 113 c 147 g 182 t
ORIGIN	
Query Match	31.3%, Score 290; DB 62; Length 620;
Best Local Similarity	72.2%; Pred. No. 4.2e-78;
Matches 377; Conservative	0; Mismatches 145; Indels 0; Gaps 0;
406	aacgccaactgtcctgtccatctctcaacctgccacagctgaagatagcttagaac 465
620	AATGCAATCTTCCAGCTCCATCTTTCACATCTTCACACAACTTAAAGCCAGCTTCAGAAAT 561
466	gttgatctgaatcgtcgtgaagccttctgtctgtctgcgagagacacattggaag 525
560	gttgatctgaatcgtcgtgaagccttctgtctgtctgcgagagacacattggaag 501
526	aaccagtgtaagttcatcatatgatatgtagtcttacaattccagcaacacactgggtttacctgac 585
500	AATCAATGTCAAGTTTATCTTCTGACAGATTATACAACTTTCACCAACACAGGTTTACCGAC 441
586	cccaagctggaacacactgaagctatctcagaacactgaagagctgtgcccaactgaatggacac 645
440	CTTCACTCACTCAACACTTACTTACCTCCAAACTCTTCTGTGACTATGCGCCCTTAATGGCAAT 381
646	ctcagtgacacagtggaatttgatctgtggggcccccaacacactcttgatataaagttact 705
380	CGAAGTGGCTTGGTGAATGATTTGATCTACGTACAGCGCTTACGGTTTTCGACACAAATACTTAC 321
706	gtgaatctagagagagcaagaagcctgtatagagatgataagaactgttttagcagttca 765
320	GTGATATCTCAAAAGAGCCGAAAGGTTCTTATCCAGAGCCGACCAAGAGTTGTCTCTAGCCCC 261
766	gagcccaactgacacatctccactgttggaagattttgttaactctaccctcaactctctt 825
260	AATGCACTGACACAAATCCCTTGTGTGAGACATATGCTGATGCGACCAACAACTTCTTC 201
826	aacgctcttcgagagacatagacgttattgttgtaactctaccctctcagtgatgccaca 885
200	AATGATTTTGTGGAGGCAATGAATGATGGGAAACATTAACCAACACTACAGAACTCAA 141
886	ggccagatctgtctgaactcagagtggtgtaaacagcaactct 927
140	GGACAAATCAGATTGAACCTGATGATTTGTGTAACCTCAACTCT 99

RESULT	2	LOCUS	AV442242	516 bp	mRNA	EST	19-Apr-2000
DEFINITION			AV442242	Arabidopsis thaliana above-ground organ two to six-week old	Arabidopsis thaliana cDNA clone AP237e04_r 5', mRNA sequence.		
ACCESSION			AV442242				
VERSION			AV442242.1	GI:7612647			
KEYWORDS			EST.				
SOURCE			thale cress.				
ORGANISM			Arabidopsis thaliana				
REFERENCE			Eukaryota: Viridiplantae; Embryophyta: Tracheophyta; Spermatophyta: Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.				
AUTHORS			1 (bases 1 to 516)				
TITLE			Nakamura, Y. and Asamizu, E.				
JOURNAL			Arabidopsis thaliana expressed sequence tags				
COMMENT			Unpublished (2000)				
FEATURES			Contact: Erika Asamizu				
SOURCE			The First Laboratory for Plant Gene Research				
			Kazusa DNA Research Institute				
			Yama 1532-3, Kisarazu, Chiba 292-0812, Japan				
			Email: asamizuekazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.				
			Location/Qualifiers				
			1..516				
			/organism="Arabidopsis thaliana"				
			/strain="Columbia"				
			/db_xref="taxon:3702"				
			/clone="AP237e04_r"				
			/clone_lib="Arabidopsis thaliana above-ground organ two to six-week old"				
			/tissue_type="above-ground organ"				
			/dev_stage="two to six-week old"				
BASE COUNT			138 a	129 c	107 g	142 t	
ORIGIN							
Query Match			29.8%,	Score 276.6;	DB 62;	Length 516;	
Best Local Similarity			71.1%;	pred. No. 5e-74;			
Matches 366;			Conservative 1;	Mismatches 149;	Indels 0;	Gaps 0;	
OY	169	gacacacacacagcttcgcacatcgaagaatgatctcggaaagctacacagcgccag	228				
DB	2	GACAAACAGACATCATTTGGAACAGAAAGATGGCTTGGAAAGCAAAATTCGGCTCGG	61				
OY	229	gaccttcagtgatcatcgacgcagcagaagcgcttgatgcagcagctgcacagcaagctc	288				
DB	62	GGATTTCCAGTATGATGATGATGAAAGTGAAGCTGGGGTGGAGAGGGCATTCGCCAAGAACCGTT	121				
OY	289	agttgtgcagacatgctgaactaagctcgcacacagagcgltgactcttgcaagcgagccg	348				
DB	122	TCATCGCAGATATATGCTCACCATTCAGCTCAACAATGCTGACATTTGGCAGAGAGTCTCT	181				
OY	349	tcctcgagagatgctcgctcggtgcgaagctgaactccctacagcagcttcctagatctggc	408				
DB	182	TCTTGGAGGGTTCCTTTGGGAAGGAGACATTTACAAAGCATTCCTGGAACTCGCTAAT	241				
OY	409	gccaaactggcctggctccacatcttgcacccgcgcgcagcaggaagatagactttrgaag	468				
DB	242	GCAAAATCTTTCCAGCTCCATTTCTTTACACTTCCACAACTTAAAGCCAGCTTCAGAAATGTT	301				
OY	469	ggctcgaacatgcgcagtgagcttggtgctctgtccgagagacacacatttgaagaagac	528				
DB	302	GGTCTCGATCGTCCTTCGATGCTCGTTCGTCCTCCGGTGGGTCACACAAATTTGGTAAAT	361				
OY	529	cagtgatgattcatcatgagatagctcttacaattccagcaacactgggtgtacctgacccc	588				
DB	362	CATATGTGATTTATTTCTTGACAGATTATACAAATTTACAGAAACACAGGTTTACCCACCT	421				
OY	589	acggcgaacactgattctcagacacactggaagcgctgtggtccactgaatgcaacctc	648				
DB	422	ACACTCAACACTACTTACCTCCAAACTCTTCTGTCGACATATGCCCTTTAATGGCAATGCA	481				
OY	649	agtgacatgagtgactttagatctgcggaaccccaac	683				

Db	482	AGTGCCTTGATGATTTTGCATCTACGTACGCTTAC	516
RESULT	3		
LOCUS	CNS00MWO/c		
DEFINITION	CNS00MWO 720 bp DNA	GSS	28-JUN-1999
ACCESSION	Arabidopsis thaliana genome survey sequence T7 end of BAC T12L18 of		
VERSION	TA00 library from strain Columbia of Arabidopsis thaliana, genomic		
KEYWORDS	survey sequence.		
SOURCE	AL094170		
ORGANISM	AL094170.1 GI:5295324		
REFERENCE	GSS.		
AUTHORS	thale cress.		
JOURNAL	Arabidopsis thaliana		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
AUTHORS	euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core		
JOURNAL	eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;		
REFERENCE	Arabidopsis.		
AUTHORS	1 (bases 1 to 720)		
JOURNAL	Salanoubat,M., Choisine,N., Artiguenave,F., Brottier,P., Wincker,P.,		
REFERENCE	Sanson,D., Saurin,W., Weissenbach,J. and Quetier,F.		
AUTHORS	Unpublished		
JOURNAL	2 (bases 1 to 720)		
REFERENCE	Genoscope.		
AUTHORS	Direct Submission		
JOURNAL	Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :		
REFERENCE	Bp 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr		
AUTHORS	- Web : www.genoscope.cns.fr)		
JOURNAL	Location/Qualifiers		
FEATURES	1..720		
SOURCE	/organism="Arabidopsis thaliana"		
	/strain="Columbia"		
	/db_xref="taxon:3702"		
	/clone_lib="TAMU"		
	/clone="T12L18"		
	/note="end : 77"		
BASE COUNT	229 a 146 c 134 g 211 t		
ORIGIN			
Query Match	25.5% Score 236.4; DB 122; Length 720;		
Best Local Similarity	71.9% Pred. No.1,2e-61;		
Matches	309; Conservative 0; Mismatches 121; Indels 0; Gaps 0;		
QY	498 tctgtcggagagacacacatttggaagaagaccagtgtagttcatcatcgaatgagctcta	557	
DB	644 TTTGTTAGGGCTCACACATTGGTAAATCAATGTCATTCATATATGAGCAGATTATA	585	
QY	558 caatttcagcaacactcgggtttaccctgagaccaccagcttgaacacttaagtatctccagaact	617	
DB	584 CAACATTACCAACACACGATCTACCTGACCTACACTCACTCAACACTACTACTCCAAACTCT	525	
QY	618 gaggagcttgcgccacatgycgaactcgaactcgaactgagcactgagcattgcatctcgagac	677	
DB	524 TCGTGTCAATGTCTCCGCAATGTGTRATAAAGCGCTTAGTGAGATTGATCTCGGTAC	465	
QY	678 cccaacacatcttcogaabaacactatgtaactctagaagaagcagaagcctgatata	737	
DB	464 GCCTTTGGTTTGCACAAATAATCTATGTGAATCTTAAAGGCAAAAAGGCTTATCCA	405	
QY	738 gaggatcagaagacigtgtagcagtcacgaagccaactgacacatccactcctggttgagaag	797	
DB	404 GAGCGACCAAGAGTGTCTTCTAGCCCAATCCCACTGACACAATCCCTTGGTGAGAC	345	
QY	798 ttctgtaactactcaaaccttcttlaagccttcgagagccttgagagccatgagacgtatagg	857	
DB	344 ATATGCTGATGGCACACAAATCTTCTCAATGCATTCTGAGAGCAATGAATAGATGGG	285	
QY	858 taacattaccctctgaagggtacccaagagccagatctgctgaactcgagatggtctaa	917	
DB	284 AATATTTACCAACACTACAGGAATCTCAAGACAAATACAGTGTGAATTTGATGAGTGTGAA	225	

oy	918	cagcaactt 927			
		1  111111			
Db	224	CTCCAACTCT 215			
RESULT	4				
LOCUS	A0956990	652 bp	DNA	GSS	28-JAN-2000
DEFINITION	LERAN24TR LERR Arabidopsis thaliana genomic clone LERAN24,				
ACCESSION	A0956990				
VERSION	A0956990.1	GI:6784691			
KEYWORDS	GSS.				
SOURCE	thale cress.				
ORGANISM	Arabidopsis thaliana				
	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta				
	Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;				
	Brassicaceae; Arabidopsids.				
	1 (bases 1 to 652)				
REFERENCE	Buell,C.R., Lin,X., Pal,G., Barnstead,M., Bowman,C., Utterbach,T.,				
AUTHORS	Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.				
	Genomic survey sequencing of Landsberg erecta ecotype of				
TITLE	Arabidopsis thaliana and identification of sequence-based				
	polymorphisms				
	Unpublished (2000)				
JOURNAL	On Dec 15, 1999 this sequence version replaced gi:4575946.				
COMMENT	Contact: Xiaoying Lin				
	The Institute for Genomic Research				
	9712 Medical Center Dr., Rockville, MD 20850, USA				
	Tel: 301 838 0200				
	Fax: 301 838 0208				
	Email: atel@tigr.org				
	For additional information, see <a href="http://www.tigr.org/tdb/at/at.html">http://www.tigr.org/tdb/at/at.html</a>				
	Seq primer: TR				
	Class: shotgun.				
FEATURES	Location/Qualifiers				
source	1..652				
	/organism="Arabidopsis thaliana"				
	/strain="Landsberg erecta"				
	/db_xref="taxon:3702"				
	/clone="LERAN24"				
	/clone_1lb="LEA"				
	/note="Organ: Leaf; Vector: PHOS1. Total genomic DNA was				
	sheared to 0.9-1 kbp before ligation."				
BASE COUNT	190 a 145 c 121 g 196 t				
ORIGIN					

QY	798	ttttcttactctactctaaactctttaaagcctctgtgaagccatgacgcgatgagg	857
Db	488	ATATGCTGATGGCACACAAACATTTCTTCAATTCATCTGTGGAGGCAATGAATGATGAGG	547
QY	858	taacatlaacccctctgaaggggtacccaaggcagaattcgtcgaactgcagagtgatca	917
Db	548	AAATTTTACACCAACACTACAGGAACTCAAGGACCAATCAAGTTGATTTGAGAGTGTGCA	607
QY	918	cagcaactct 927	
Db	608	CTCCACATCT 617	
RESULT	5		
LOCUS	AA042293		
DEFINITION	24630 CD4-13 Arabidopsis thaliana cDNA clone E6A9r7, mRNA sequence.		
ACCESSION	AA042293		
VERSION	AA042293.1	GI:2414082	
KEYWORDS	EST.		
SOURCE	thale cress.		
ORGANISM	Arabidopsis thaliana		
REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		
AUTHORS	1 (bases 1 to 564) Newman,T., deBruin,J.F., Green,P., Keegstra,K., Kende,H., McIntosh,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M., Retzel,E. and Somerville,C.		
TITLE	Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones		
JOURNAL	Plant Physiol. 106, 1241-1255 (1994)		
MEDLINE	95148729		
COMMENT	On Sep 19, 1997 this sequence version replaced gi:1520451.		

## FEATURES

### source

[illegible]

*note="vector: pluscript sr"; site.1:ecor1; site.2: ecor1; using 5 ug of polyadenylated mRNA from 3 day-old Arabidopsis thaliana (Columbia) seedling hypocotyls as template and oligo d(T) as primer, first strand synthesis was catalyzed by Moloney murine leukemia virus reverse transcriptase (Pharmacia). Second-strand cDNA was made using the procedure of Gubler and Hoffman (1983) except that DNA ligase was omitted. After the second strand reaction, the ends of the cDNA were made blunt with Klenow fragment and EcoRI/NotI adapters (pharmacia) were ligated to each end. The cDNA was purified from unligated adapters by spun column chromatography using sepharyl 3-500 and size-fractionated on a 1% low melting point mini-gel. Size selected cDNAs (0.5 - 1 kb) were removed from the gel using agarase (New England Biolabs), phenol:chloroform extracted and precipitated using 0.3 M NaOAc (pH 7)/ethanol. A portion of each cDNA size-fraction (0.1 ug) was co-precipitated with 1 ug of lambdaZapIII (Stratagene/EcoRI digested, dephosphorylated arms and then ligated in a volume of 4 ul overnight. Each ligation mix was packaged in vitro using GigaPack II gold*



Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.

# REFERENCE

1 (bases 1 to 824)

# AUTHORS

Shoenaker, R., Keim, P., Vodka, L., Erpelting, J., Coryell, V., Khanna, R., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stepien, M., Thelning, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterson, R. and Wilson, R.

# TITLE

Public Soybean EST Project

# JOURNAL

Unpublished (1999)

# COMMENT

On Jan 19, 1998 this sequence version replaced gi:2284894.  
Contact: Shoenaker R./Public Soybean EST Project  
Public Soybean EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu  
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com  
Possible reversed clone: similarity on wrong strand  
High quality sequence stop: 424.

# FEATURES

source

1..824  
/organism="Glycine max"  
/db\_xref="taxon:3847"  
/clone="GENOME SYSTEMS CLONE ID: Gm-c1019-1046"  
/clone\_lib="Gm-c1019"  
/tissue\_type="immature seed coats of greenhouse grown plants"  
/lab\_host="DH10B (Gibco BRL)"  
/note="Vector: pSPORT1 (Life Technologies); Site\_1: Not I; Site\_2: Sal I; This cDNA library was constructed from mRNA isolated from immature seed coats (200-300 mgs) of greenhouse grown plants. The library was prepared using the Life Technologies superscript cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly (dT) sequence with a Not I restriction site. Sal I linker adapters were ligated to the blunt-ended cDNA fragments followed by Not I digestion. The cDNA fragments were directionally cloned into the Not I-Sal I restriction site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E.coli ElectroMax DH10B host cells (Gibco BRL). This library was constructed by Dr. Lila Vodka and Dr. Anu Khanna."

BASE COUNT 220 a 198 c 165 g 236 t 5 others  
ORIGIN

Query Match 22.0%; Score 204.4; DB 46; Length 824;  
Best Local Similarity 55.3%; Pred. No. 8.1e-52;

Matches 413; Conservative 0; Mismatches 331; Indels 3; Gaps 1;

4 caattaccctcatctacgacgaatagctgtcccaagctgtcccaactgttcgagac 63

81 CACCTTACTCTCTCGTTCTACAGAGAACATGTCCAAATGTGCTTATGTTGGA 140

64 acaatcgtaacagagctcagatccgatacagagatcgctcgaatattacgtcag 123

141 GATATCTTCGATGCTTCTTTCACCGAGATCCCGAATCGGCGCATCTCATGAGGCTTCA 200

124 ttccatgactgtctcgtaagtgctgcagcagtagacataltactggaacaacaccagt 183

201 TTTCATGATGATGCTTTGTCACAGCTTGATGATCACTTTGCTGACACACATGATACA 260

184 ttccgacatgaagaagatgattcgcggaacgctaacagcgccggggtcttcagtgatc 243

261 ATAGAAAGCGAGCAAGTGCATTCACAAATATCAATCAATTAAGAGATTGAGCGTTGTC 320

244 gatccatgaagagctgcgcttgagtcagcatgcccacgaacagtcagttgagagacctg 303

321 AATGACATCAAGACAGCGGGTGGAAATAGTTGTCCACACACAGTTTCTGTGATATT 380

304 ctgactatagctgcgaacaacagagctgactctgcagcgagaccttcgtgagagtgccg 363

381 CTTGCTTTTCACCTCACTCACTCACTTAAAGCTTCTTCTTCTTCAAGTCTCAA---CACC 440

364 ctggttcgagtgactccctcaacagagcttctctgattctgcgaacgcaactgtctgt 423

441 TTAGGAAGAGGACAGCTTACACGAAACCGAACCCTTCAATCAAAACCTTCAGCA 500

424 ccatctccaccctcccccagctgaagatagcttgaagacgtggctgcatgctgcg 483

501 CTTTCTTCAACTCACTCACTCACTTAAAGCTTCTTCTTCTTCAAGTCTCAA---CACC 557

484 agtgaacctgtgctctgtccggaagacacacattggaagaagaccagtgatgctc 543

558 CTTGATTTAGTTACACTCTCANGTGTGATAGCGTNGAGAGAGCTGCGTGCATACATTTC 617

544 atgagatagctcacaatttcacgaacacagctggttaacctcaacccagctgagacctagc 603

618 ATANACGATTATACACTTACCTACCTGGAACCCCTGATCCACTGGAACACTACA 677

604 tatccgaagacactgagagagctgtgtcccactgattgcaacactcagtgactagtgagac 663

678 TACTTTAGATTAATTTGCGTGTGATAGTCCCGACAGATGCGCATGTATNACCTCACCAAT 737

664 ttgattctgcgagaccccaacacactctcgatlaacaagtaactatgtgaactagagagcag 723

738 NTGAGACTGAGACACCTGATCAATTTGACACAGATACTACTCCATCTTCTGCAGCTT 797

724 aaagagctgtacagagtgatgaagaa 750

798 AATGGCTACCTCAGAGTGACCAAGAA 824

RESULT 8  
LOCUS T88184 499 bp mRNA EST 09-JAN-1998  
T88184  
DEFINITION 11880 Lambda-PRL2 Arabidopsis thaliana cDNA clone 157E2277, mRNA sequence.

ACCESSION T88184  
VERSION T88184.1 GI:36039

KEYWORDS EST.

SOURCE thale cress.

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 499)

REFERENCE Newman, T., deBruijn, F.J., Green, P., Keegstra, K., Kende, H., McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomas, M., Retzel, E. and Somerville, C.

Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones Plant Physiol. 106, 1241-1255 (1994)

95148729

COMMENT On Apr 14, 1993 this sequence version replaced gi:716536.

Contact: Thomas Newman

MSU-DOE Plant Research Laboratory

Michigan State University

MSU-DOE-PR, Michigan State University, Plant Biology Bldg., E. Lansing, MI

Tel: 517-353-0854

Fax: 517-353-9168

Email: 22313tcn@ldm.cl.msu.edu

Seq primer: T7 dye primer.

Location/Qualifiers

1..499

/organism="Arabidopsis thaliana"

/strain="var columbia"







KEYWORDS	EST
SOURCE	Zea mays.
ORGANISM	Zea mays.
REFERENCE	Eukaryotes: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
AUTHORS	1 (bases 1 to 625)
TITLE	Maize ESTs from various cDNA libraries sequenced at Stanford University
JOURNAL	Unpublished (1999)
COMMENT	On May 7, 1998 this sequence version replaced gi:3119571.
	Contact: Walbot V
	Department of Biological Sciences
	Stanford University
	855 California Ave, Palo Alto, CA 94304, USA
	Tel: 650 723 2227
	Fax: 650 725 8221
	Email: walbot@stanford.edu
	Plate: 707014 row: F column: 11.
FEATURES	Location/Qualifiers
source	1..625
	/organism="Zea mays"
	/cultivar="W23"
	/db_xref="taxon:4577"
	/clone_lib="707 - Mixed adult tissues from Walbot lab (SK)"
	/tissue_type="tassel, kernel, silk, husk, root, leaf"
	/dev_stage="adult"
	/lab_host="DH10B"
	/note="Organ: tassel, kernel, silk, husk, root, leaf; Vector: pAD10; Site:1: EcoRI; cDNA library from fully differentiated maize tissues from an active Mutator plant. Tissue ratio is 4/2/1/1/1 (tassel, kernel, silk, husk, root, leaf). Unidirectionally cloned."
BASE COUNT	131 a 181 c 175 g 138 t
ORIGIN	
Query Match	18.6%; Score 172.4; DB 72; Length 625;
Best Local Similarity	56.9%; Pred. No. 4.7e-42;
Matches 336; Conservative	0; Mismatches 251; Indels 3; Gaps 1;
OY	7 ttaaccctacattctacgacaatagctgtgccaaagcgtgtccaaacatgcgttcgcgacaca 66
Db	28 TTGTTCACACAGCTTCTATGACCATTCGTGCCCCAAGCGAAGAGATTTGTGCAGATTCATT 87
OY	67 atcgtcaacgagctcagatccgatcccgaggttcgtgtcttaatttaactgtctgaactc 126
Db	88 GTGGCACAGGCTGTGCGCCAGAGAGACACGAGATGGCGCATCTTATGTCAGATGAGATTTC 147
OY	127 catgactccttcgttaatggtgtgcgaagctcagcatattacttgagacaacacacagtttc 186
Db	148 CATGACTCCTTTGTCAAGGCGCTGCCATGCTTTCGTGCTGTTCGACAAACAGCAGCAGCAT 207
OY	187 cgcactgaaagatgatcctcggaagcgtlaacagcgccaggggctttccagtatcat 246
Db	208 GTTAGTGAAGAAAGGCTCCACACCGACAGCAAGAACTCCCTACAGGCGCTTTCAGCTGATFCGAC 267
OY	247 cgcataagaggttcgcgttgagtcaagtcagccacgaagaacatgactgtgtgcagactgctg 306
Db	268 CAGATTAAAGGCTGCTTGTAGGCTGCTGCGCCAGGACACAGTCTCTGTGCGCACATTGT 327
OY	307 actatagctgcgaacagagcgtgactctcttcgagcgagcgcgtcctctgagaggtgcgcctc 366
Db	328 GCCCTTGGGCTCGTGATTTCCACGCCCTCGTTGGTGTGCACTACTGGAGCTGCGCATT 387
OY	367 gttcgaagtgactccctacaggaatcttcctagatctgtgcgaacgccaactgtcctgtcca 426
Db	388 GCGCGGAGAGACTCGCTCGGTGTCAGAGCAATCCAGGCGCTCCAAACATGACATCCACGCCCC 447
OY	427 ttcttcaccctggcccgagctgaagatagctttaaagcgtgtggtctgtatctgcctcaggt 486
Db	448 AACCAACACTCCCACTATCATCAACCAAGTCAAGCGCCAGGCGCTCAATGTGT -T 504

	QY	547	gataggcttacaatttcagcaacactlgttgatcagaccacgcagtga 596   +
	Dn	565	CAGAGGCTATACACCAAGCAGCAATGGCATGGCTGACAGCACACTGGA 614
	RESULT	13	
	LOCUS	AM278775	
	DEFINITION	AM278775 691 bp mRNA EST 18-APR-2000	
	ACCESSION	Gm-c1019-4060.5 similar to TR:022443 022443 SRED COAT PEROXIDASE	
	VERSION	AM278775	
	KEYWORDS	PBCUNSORO , mRNA sequence.	
	SOURCE	AM278775 GI:6667324 EST.	
	ORGANISM	soybean. Glycine max	
	REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycyne. 1 (bases 1 to 691)	
	AUTHORS	Shoemaker,R., Keim,P., Vocklin,L., Erpelting,J., Correll,V., Khanna,A., Bollé,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Page,D., Harvey,N., Schurt,R., Rlter,E., Kohn,S., Shln,T., Jackson,Y., Cardenas,M., Mcann,R., Waterston,R. and Wilson,R. Public Soybean EST Project Unpublished (1999) On Nov. 22, 1999 this sequence version replaced gi:6461883. Contact: Shoemaker R/Public Soybean EST project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel.: 314 286 1800 Fax: 314 286 1810 Email: estewatson.wustl.edu This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3322 FAX:(888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com Insert Length: 1137 Std Error: 0.00 Seq primer: -40RP from Glibco High quality sequence stop: 446. Location/Qualifiers 1..691 /organism="Glycine max" /db_xref="taxon:3847" /clone="GENOME SYSTEMS CLONE ID: Gm-c1019-4060" /clone_1fb="Gm-c1019" /issue_type="Immature seed coats of greenhouse grown plants" /lab_host="DH10B (Gibco BRL)" /note="Vector: pSPORI1 (Life Technologies); Site_1: Not I Site_2: Sal I; This cDNA library was constructed from mRNA isolated from immature seed coats (200-300 mgs) of greenhouse grown plants. The library was prepared using the life technologies psupertscript cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly (dt) sequence with a Not I restriction site. Sal I linker adapters were ligated to the blunt-ended cDNA fragments followed by Not I digestion. The cDNA fragments were directionally cloned into the Not I-Sal I restriction site of the pSPORI1 vector. The ligated cDNA fragments were transformed into E.coli Electromax DH10B host cells (Gibco BRL). This library was constructed by Dr. Lila Vocklin and Dr. Anu Khanna."	
	TITLE	JOURNAL	
	COMMENT		
	BASE COUNT	208 a 168 c 131 g 182 t 2 others	





---



Db 121 PIGRRSLQAFLDLANANLPAFFETLPQKDSFRNGLNRSDDLVALSGHTEGKNOCRF 180  
181 IMRLVNFSTNGLPDPLNTNTYLOTLRGLCPNLGNLSALVDFDLRTPTIDNKYYVLEE 240  
Db 181 IMRLVNFSTNGLPDPLNTNTYLOTLRGLCPNLGNLSALVDFDLRTPTIDNKYYVLEE 240  
QY 241 QKGLIQSDQELFSSPDADTITPLVRSFANSTQTFENAFVEMDMGNITPLTGTGQIRL 300  
Db 241 QKGLIQSDQELFSSPDADTITPLVRSFANSTQTFENAFVEMDMGNITPLTGTGQIRL 300  
QY 301 NCRVNSNS 309  
Db 301 NCRVNSNS 309

RESULT 2  
P81167  
ID P81167 standard: protein; 308 AA.  
AC P81167;  
DT 05-NOV-1990 (first entry)  
DE Horseradish peroxidase.  
KW horseradish peroxidase; Armoracia lapathifolia;  
KM Armoracia rusticana; ss.  
OS Armoracia sp.  
PN J63207386-A.  
PD 26-AUG-1988.  
PF 19-FEB-1987; 037623.  
PR 19-FEB-1987; JP-037623.  
PA (TOYO) Toyobo KK.  
DR N-PSDB; N81308.  
PT New structural gene of horseradish peroxidase -  
PT produced under artificially controlled conditions by gene  
PT manipulation.  
PS Disclosure: P: Japanese.  
CC DNA was preferably isolated from tissue cultures of root cells of  
CC Armoracia sp. The gene can be inserted into a plasmid for  
CC transformation of suitable host cells. The product (HRP) can then  
CC be collected from the culture medium.  
CC See also N82195.  
SQ Sequence 308 AA;

Query Match 99.4%; Score 1598; DB 1; Length 308;  
Best Local Similarity 99.7%; Pred. No. 3.7e-155;  
Matches 307; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 QLTPTFYDNCSPVNSIVRDTIVNELRSDPRIAASILRLHFHDCFVNGCDASILLNNTTS 61  
Db 1 QLTPTFYDNCSPVNSIVRDTIVNELRSDPRIAASILRLHFHDCFVNGCDASILLNNTTS 60  
QY 62 FRTKEKDAFGANARSARGFVYIDRMAKAVERACPTVSCADLLTTAAOOSVTLAGPSMRVP 121  
Db 61 FRTKEKDAFGANARSARGFVYIDRMAKAVERACPTVSCADLLTTAAOOSVTLAGPSMRVP 120  
QY 122 LGRDRLQAFLDLANANLPAFFETLPQKDSFRNGLNRSDDLVALSGHTEGKNOCRF 181  
Db 121 LGRDRLQAFLDLANANLPAFFETLPQKDSFRNGLNRSDDLVALSGHTEGKNOCRF 180  
QY 182 MDRLVNFSTNGLPDPLNTNTYLOTLRGLCPNLGNLSALVDFDLRTPTIDNKYYVLEE 241  
Db 181 MDRLVNFSTNGLPDPLNTNTYLOTLRGLCPNLGNLSALVDFDLRTPTIDNKYYVLEE 240  
QY 242 KGLIQSDQELFSSPDADTITPLVRSFANSTQTFENAFVEMDMGNITPLTGTGQIRL 301  
Db 241 KGLIQSDQELFSSPDADTITPLVRSFANSTQTFENAFVEMDMGNITPLTGTGQIRL 300  
QY 302 CRVNSNS 309  
Db 301 CRVNSNS 308

RESULT 3  
ID P95203  
AC P95203;  
DT 18-JUL-1990 (first entry)  
DE Horseradish peroxidase (HRP) as encoded by HRP gene modified for  
DE expression in mammalian cells  
KW Horseradish peroxidase; restriction sites; DNA construct;  
KW fusion protein; enzyme assays; immunoassays; synthetic gene;  
KM synthetic linker; pSD18.  
OS Horseradish.

FH Key Location/Qualifiers  
FT peptide 1..19  
FT protein /note="N-terminal pre sequence"  
FT peptide 20..327  
FT peptide 328..342  
FT peptide /note="C-terminal pro sequence"  
PN MO8903424-A.  
PD 20-APR-1989.  
PF 07-OCT-1988; G00833.  
PR 08-OCT-1987; GA-023662.  
PA (BRRI-) Brit Bio-Tech Ltd.  
PI Edwards RM, Burke JF;  
DR WPI; 89-130042/17.  
DR N-PSDB; N90762.  
PT Synthetic gene encoding horseradish peroxidase -  
PT used for efficient produ. of the enzyme or fusion prods. for  
PT use in enzyme assays, esp. immunoassays  
PS Disclosure: 45pp; English.  
CC The synthetic gene is prep'd. by coupling successive nucleotides and/or  
CC ligating appropriate oligomers. A construct containing the DNA which  
CC encodes the mature protein (see FT) is claimed. The construct can be a  
CC vector where the HRP gene and a desired gene are co-expressed as a single  
CC fusion prod., a single polycistronic message, or 2 separate but linked  
CC transcriptional units. The fusion proteins are for use in enzyme assays,  
CC esp. immunoassays. Genes carrying mutations that destroy HRP activity  
CC can be used to follow the frequency of reversion or suppression of  
CC the particular mutation introduced into the gene. The constructs can  
CC be used in transgenic animals, esp. mammals.  
SQ Sequence 342 AA;

Query Match 99.4%; Score 1598; DB 1; Length 342;  
Best Local Similarity 99.7%; Pred. No. 4.3e-155;  
Matches 307; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 QLTPTFYDNCSPVNSIVRDTIVNELRSDPRIAASILRLHFHDCFVNGCDASILLNNTTS 61  
Db 20 QLTPTFYDNCSPVNSIVRDTIVNELRSDPRIAASILRLHFHDCFVNGCDASILLNNTTS 79  
QY 62 FRTKEKDAFGANARSARGFVYIDRMAKAVERACPTVSCADLLTTAAOOSVTLAGPSMRVP 121  
Db 80 FRTKEKDAFGANARSARGFVYIDRMAKAVERACPTVSCADLLTTAAOOSVTLAGPSMRVP 139  
QY 122 LGRDRLQAFLDLANANLPAFFETLPQKDSFRNGLNRSDDLVALSGHTEGKNOCRF 181  
Db 140 LGRDRLQAFLDLANANLPAFFETLPQKDSFRNGLNRSDDLVALSGHTEGKNOCRF 199  
QY 182 MDRLVNFSTNGLPDPLNTNTYLOTLRGLCPNLGNLSALVDFDLRTPTIDNKYYVLEE 241  
Db 200 MDRLVNFSTNGLPDPLNTNTYLOTLRGLCPNLGNLSALVDFDLRTPTIDNKYYVLEE 259  
QY 242 KGLIQSDQELFSSPDADTITPLVRSFANSTQTFENAFVEMDMGNITPLTGTGQIRL 301  
Db 260 KGLIQSDQELFSSPDADTITPLVRSFANSTQTFENAFVEMDMGNITPLTGTGQIRL 319  
QY 302 CRVNSNS 309  
Db 320 CRVNSNS 327

RESULT 4  
P94810



ID P94810 standard; protein; 308 AA.  
AC P94810;  
DT 10-MAR-1993 (revised)  
DE 25-JUN-1989 (first entry)  
KW Horseradish peroxidase (HRP) Isoenzyme C; detection reagent;  
FH superoxide radicals (SOD); ss.  
FT Key Location/Qualifiers  
FT cds 1..924  
PN EP-299682-A.  
PD 18-JAN-1989.  
PF 07-JUL-1988; 306222.  
PR 17-JUL-1987; GB-016938.  
PA (AMER-) Amersham Int. PLC.  
PI Chiswell DJ, Orlepp SA;  
DR WPI: 89-017184/03.  
DR N-PSDB: N93078.  
PT New synthetic DNA encoding horseradish peroxidase C Isoenzyme or variants  
PT or active fragments, useful eg as detection reagent.  
PS Disclosure; fig.6; 19pp; English.  
CC This sequence contains the horseradish peroxidase Isoenzyme C (HRP-C)  
CC complete sequence (309 AA's) and flanking AA's. Production is by recomb-  
CC inant techniques. A conservative variant can also be encoded by the  
CC corresponding DNA sequence by introducing base changes resulting in amino  
CC acid (AA) substitutions. Properties of this enzyme are essentially the  
CC same as the native HRP-C. A peptide contg. at least one HRP-C epitope can  
CC also be encoded by restriction enzyme cleavage of the same DNA sequence  
CC resulting in peptides of different lengths (contg. the active AA's Arg-183  
CC and Tyr-185). HRP-C is a scavenger of toxic superoxide radicals (SOD) and  
CC is useful as a detection reagent. It provides enhanced illumination by  
CC catalysis of a substrate such as luminol, in presence of a light-output  
CC enhancer.  
SO Sequence 308 AA;

Query Match 98.8%; Score 1589; DB 1; Length 308;  
Best Local Similarity 99.4%; Pred. No. 3e-154;  
Matches 306; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 QLTPEFYNSCPNVSNTVDRITVNELRSDPRIASILRLHFHDFVNGCASIILDDNTTS 61  
DB 1 QLTPEFYNSCPNVSNTVDRITVNELRSDPRIASILRLHFHDFVNGCASIILDDNTTS 60  
QY 62 FRTEKDAFGNANSARGFPVIDRMKAIVESACPRIVSCADLLTTIAQOSVTLAGGSPMRV 121  
DB 61 FRTEKDAFGNANSARGFPVIDRMKAIVESACPRIVSCADLLTTIAQOSVTLAGGSPMRV 120  
QY 122 LGRDLSQAFLDLANANLPAFFFTLPOLKDSFRNVLNRSSDLVALSGHTEGKNOCRF 181  
DB 121 LGRDLSQAFLDLANANLPAFFFTLPOLKDSFRNVLNRSSDLVALSGHTEGKNOCRF 180  
QY 182 MDRLYNFSNTGLPDPPTLNTTYLQTLRGICPLNGLSALVDFDLTPRTTFDNKYYVNEEQ 241  
DB 181 MDRLYNFSNTGLPDPPTLNTTYLQTLRGICPLNGLSALVDFDLTPRTTFDNKYYVNEEQ 240  
QY 242 KGLIOSDDELFPSPDATDTIPLVRSFANSTOTFFNAFEADRMGNITPLATOGQIRLN 301  
DB 241 KGLIOSDDELFPSPDATDTIPLVRSFANSTOTFFNAFEADRMGNITPLATOGQIRLN 300  
QY 302 CRVNSNS 309  
DB 301 CRVNSNS 308  
RESULT 5  
ID R32760  
AC R32760;  
DT 23-JUN-1993 (first entry)  
DE HRP/IGF-II fusion protein.  
KW Insulin receptor; E.coli periplasmic protease resistance;  
KW wound healing; osteoporosis; muscle-wasting; growth disorder;  
KW horseradish peroxidase; inclusion bodies; ds.

OS Synthetic.  
FH Key Location/Qualifiers  
FT Region 1..282  
FT /note="first 282 residues of HRP"  
FT protein 283..349  
FT /label= IGF-II  
PN M09303152-A.  
PD 18-FEB-1993.  
PF 27-JUL-1992; G01389.  
PR 29-JUL-1991; GB-016325.  
PR 05-FEB-1992; GB-002401.  
PA (BRBI-) BRITISH BIO-TECHNOLOGY LTD.  
PI Bawden L, Edwards RM;  
DR WPI: 93-076509/09.  
DR N-PSDB: 037614.  
PT New insulin-like Growth Factor II analogues - for treating growth  
PT disorders, osteoporosis and muscle wasting diseases, have less  
PT affinity for insulin receptors than unmodified IGF-II  
PS Example 4; Page 44-46; 71pp; English.  
CC This fusion protein comprises the first 282 amino acids of HRP  
CC fused to IGF-II via a methionine residue to allow for the release  
CC of mature IGF-II by cleavage with CNBr. The fusion lacks the final  
CC 27 amino acids of HRP. This fusion protein was not ideal for  
CC production of IGF-II; the IGF-II is only 19% of the total fusion  
CC protein and the HRP sequence contains two internal methionine  
CC residues which act as additional cleavage sites. The IGF-II is the  
CC wild-type sequence, i.e. with Arg residues at positions 37 and 38.  
CC The invention covers analogues of this sequence in which one or both  
CC of the Arg residues are subst. by other amino acids (pref. non-  
CC basic and esp. Gln). The mutants are protease-resistant and have  
CC reduced affinity for insulin receptors.  
SO Sequence 349 AA;

Query Match 90.9%; Score 1462; DB 1; Length 349;  
Best Local Similarity 99.6%; Pred. No. 3.4e-141;  
Matches 281; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLPTEFYNSCPNVSNTVDRITVNELRSDPRIASILRLHFHDFVNGCASIILDDNTT 60  
DB 1 MQLPTEFYNSCPNVSNTVDRITVNELRSDPRIASILRLHFHDFVNGCASIILDDNTT 60  
QY 61 SFRTKDAFGNANSARGFPVIDRMKAIVESACPRIVSCADLLTTIAQOSVTLAGGSPMRV 120  
DB 61 SFRTKDAFGNANSARGFPVIDRMKAIVESACPRIVSCADLLTTIAQOSVTLAGGSPMRV 120  
QY 121 PLGRDLSQAFLDLANANLPAFFFTLPOLKDSFRNVLNRSSDLVALSGHTEGKNOCRF 180  
DB 121 PLGRDLSQAFLDLANANLPAFFFTLPOLKDSFRNVLNRSSDLVALSGHTEGKNOCRF 180  
QY 181 IMDRLYNFSNTGLPDPPTLNTTYLQTLRGICPLNGLSALVDFDLTPRTTFDNKYYVNEEQ 240  
DB 181 IMDRLYNFSNTGLPDPPTLNTTYLQTLRGICPLNGLSALVDFDLTPRTTFDNKYYVNEEQ 240  
QY 241 OKGLIOSDDELFPSPDATDTIPLVRSFANSTOTFFNAFEAM 282  
DB 241 OKGLIOSDDELFPSPDATDTIPLVRSFANSTOTFFNAFEAM 282  
RESULT 6  
ID W44931  
AC W44931 standard; protein; 352 AA.  
DT 09-NOV-1998 (first entry)  
DE Soybean seed coat peroxidase.  
KW Peroxidase; seed coat; soybean; transgenic plant.  
OS Glycine max cv. Harosoy 63.  
FH Key Location/Qualifiers  
FT Peptide 1..26  
FT /label= Sig\_peptide  
FT protein 29..352  
FT /label= Mat\_protein  
FT Active\_site 59..70

FT Domain 188..197  
 FT /label= Haem-binding  
 PN CA2211018-A.  
 PD 30-MAR-1998.  
 PF 19-SEP-1997; 211018.  
 PR 30-SEP-1996; US-722414.  
 PR 30-SEP-1996; CA-186833.  
 PA (MIAC ) CANADA MIN AGRICULTURE.  
 PI Gijzen M.  
 DR WPI: 98-388720/34.  
 DR N-PSDB: V30460-61.  
 PT New DNA used in preparation of soybean seed coat peroxidase - used  
 PT e.g. in diagnostic assays, for de-linking waste paper, for oxidation  
 PT of primary alcohols, in paper and pulp industries and to encode  
 PT herbicide resistance in plants  
 PS Example 1: Page 42-45; 75pp: English.  
 CC This is the amino acid sequence of soybean seed coat peroxidase.  
 CC Its was deduced from isolated cDNA and genomic DNA clones (see  
 CC V30460-61). Vectors, host cells, and transgenic plants are  
 CC provided, as well as methods for the production of recombinant  
 CC seed coat peroxidase. Soybean coat peroxidases are useful  
 CC industrially, e.g. their high thermal stability even at extremes of  
 CC pH make them useful in diagnostic assays as an enzyme label for  
 CC e.g. antigens, and within staining techniques. They are also  
 CC useful for de-linking printed waste paper, biocatalytic oxidation  
 CC of primary alcohols, as a replacement for chlorine in the pulp and  
 CC paper industry or as a formaldehyde replacement for use in  
 CC adhesives.  
 SO Sequence 352 AA:

Query Match 57.3%; Score 921; DB 1; Length 352;  
 Best Local Similarity 59.1%; Pred. No. 5.8e-86;

Matches 182; Conservative 40; Mismatches 84; Indels 2; Gaps 2;

QY 2 QLTPTFDNSCPNVSIVRDTIVNELSDPRIASILRLHFHFCFVNGCDASILLDNWTS 61  
 DB 27 QLTPTFRETCPNLFPIVFGVIFDASFTDPRIGASLMRLHFHDFVOGCGSVLNNWTF 86  
 QY 62 FRTKEKDFGANSARGPVVDKRAAVESACPTVSCADLLTTAAQOSVTLAGPSPMRVP 121  
 DB 87 IESQDLPINISIRGLDVNDIKTAVENSCPTVSCADLLTAAEIASVYLGSGPMPVP 146  
 QY 122 LGRDLSQAFLDLANANLPAFFTLPLQKDSFRNNGNRSDDLVALSGHTFGKNOCRFI 181  
 DB 147 LGRDLSLTANRTLANONLPAFFNLTLQKASFAVOGLN-TLIDLVTLSGHTFGARCTF 205  
 QY 182 MDRLYNESNTGLPDPPLNTTYLQTLRGCLPLNGNLSALVDFDLRTPTIFDNKYYVNL 241  
 DB 206 INRLYNESNTGNPDPTLNTTYLLEVLRARCPQMATGDMLTLDLSTPDQFNRYSNLQL 265  
 QY 242 KGLISQDELFSSPDATDTPLVRSFANSQTGFENAFVEMADRMGNITPTLTGQGI 301  
 DB 266 NGLISQDELFSTPGA-DTILPIVNSFSSNQTFFSNFRVSMIKMGNIGVLTGDEGEIRLQ 324  
 QY 302 CRVYNSNS 309  
 DB 325 CNFVNGDS 332

RESULT 7  
 ID W73138 standard; Protein: 352 AA.  
 AC W73138;  
 DT 02-FEB-1999 (first entry)  
 DE Soybean seed coat peroxidase.  
 KW Peroxidase; seed coat; soybean; transgenic plant.  
 OS Glycine max cv. Harosoy 63.  
 FH Key Location/Qualifiers  
 FT Peptide 1..26  
 FT /label= Sig\_peptide  
 FT 27..352  
 FT Protein /label= Mat\_protein

FT Active\_site 59..70  
 FT Domain 188..197  
 FT /note= "haem-binding domain"  
 PN CA2186833-A.  
 PD 31-MAR-1998.  
 PF 30-SEP-1996; 186833.  
 PR 30-SEP-1996; CA-186833.  
 PA (MIAC ) CANADA MIN AGRICULTURE.  
 PI Gijzen M.  
 DR WPI: 98-52191/45.  
 DR N-PSDB: V59077-78.  
 PT New DNA encoding soya seed-coat peroxidase and the promoter - useful  
 PT for expression of heterologous genes in the seed coat, which may  
 PT give herbicide resistance, viral protection and pest and pathogen  
 PT control, and may improve the taste or nutritional value of soya  
 PT beans.  
 PS Example 1: Fig 1; 61pp: English.  
 CC This is the amino acid sequence of a soybean seed coat peroxidase,  
 CC as deduced from an isolated cDNA clone sequence (see V59077).  
 CC Soybean plants containing a dominant bp allele accumulate large  
 CC amounts of peroxidase in the hourglass cells of the subepidermis.  
 CC Homozygous recessive eep genotypes are much reduced in total seed  
 CC coat peroxidase activity. Soybean peroxidases have been shown to  
 CC have useful catalytic properties and a high degree of thermal  
 CC stability even at extremes of pH. They are used as enzyme labels  
 CC in diagnostic assays. The invention relates to soybean seed coat  
 CC polynucleotides, and to regulatory regions of the peroxidase gene  
 CC that can be used to control the expression of genes of interest  
 CC within the seed coat.  
 SO Sequence 352 AA:

Query Match 57.3%; Score 921; DB 1; Length 352;  
 Best Local Similarity 59.1%; Pred. No. 5.8e-86;

Matches 182; Conservative 40; Mismatches 84; Indels 2; Gaps 2;

QY 2 QLTPTFDNSCPNVSIVRDTIVNELSDPRIASILRLHFHFCFVNGCDASILLDNWTS 61  
 DB 27 QLTPTFRETCPNLFPIVFGVIFDASFTDPRIGASLMRLHFHDFVOGCGSVLNNWTF 86  
 QY 62 FRTKEKDFGANSARGPVVDKRAAVESACPTVSCADLLTTAAQOSVTLAGPSPMRVP 121  
 DB 87 IESQDLPINISIRGLDVNDIKTAVENSCPTVSCADLLTAAEIASVYLGSGPMPVP 146  
 QY 122 LGRDLSQAFLDLANANLPAFFTLPLQKDSFRNNGNRSDDLVALSGHTFGKNOCRFI 181  
 DB 147 LGRDLSLTANRTLANONLPAFFNLTLQKASFAVOGLN-TLIDLVTLSGHTFGARCTF 205  
 QY 182 MDRLYNESNTGLPDPPLNTTYLQTLRGCLPLNGNLSALVDFDLRTPTIFDNKYYVNL 241  
 DB 206 INRLYNESNTGNPDPTLNTTYLLEVLRARCPQMATGDMLTLDLSTPDQFNRYSNLQL 265  
 QY 242 KGLISQDELFSSPDATDTPLVRSFANSQTGFENAFVEMADRMGNITPTLTGQGI 301  
 DB 266 NGLISQDELFSTPGA-DTILPIVNSFSSNQTFFSNFRVSMIKMGNIGVLTGDEGEIRLQ 324  
 QY 302 CRVYNSNS 309  
 DB 325 CNFVNGDS 332

RESULT 8  
 ID W87893 standard; Protein: 333 AA.  
 AC W87893;  
 DT 12-APR-1999 (first entry)  
 DE Maize Per5 root preferential cationic peroxidase.  
 KW Peroxidase; Per5; maize; corn; transgenic plant; promoter;  
 OS Zea mays.  
 FH Key Location/Qualifiers  
 FT Peptide 1..32  
 FT /note= "signal peptide"

FT	Protein	33..333	
FT		/note="mature protein"	
FT	Modified_site	85	
FT		/note="putative N-glycosylation"	
FT	Modified_site	170	
FT		/note="putative N-glycosylation"	
FT	Modified_site	213	
FT		/note="putative N-glycosylation"	
FT	Modified_site	311	
FT		/note="putative N-glycosylation"	
PN	W09856921-A1.		
PD	17-DEC-1998.		
PR	10-JUN-1998;	U11921.	
PR	12-JUN-1997;	US-049752.	
PA	(DOMC ) DOM AGROSCIENCES LLC.		
PI	Ainley M, Armstrong K, Belmar S, Folkerts O, Hopkins N,		
PI	Menke M, Paredy D, Petcolino JF, Smith K, Woosley A;		
DR	WPI; 99-080904/07.		
DR	N-PDB: V63717 AND V63720.		
PT	New isolated regulatory root sequences for transgenic plants - which are derived from the maize root preferential cationic peroxidase protein (per5) gene		
PT	[disclosure: Page 89-90; 150pp: English.		
CC	This is the amino acid sequence of peroxidase Per5 of maize. The enzyme is encoded by the root preferential cationic peroxidase per5 gene (see V63717). Regulatory sequences of the per5 gene, including the promoter, introns and 3' untranslated region (3'UTR), are used in claimed recombinant gene cassettes for controlling expression of recombinant genes in selected tissue, especially the root, of transformed plants, particularly maize. The gene cassettes can be used for expression of heterologous genes such as genes that confer tolerance to herbicides, insects or viruses, and characteristics to the plant.		
CC	Sequence 333 AA:		

Query Match	47.8%	Score 768.5;	DB 1;	Length 333;
Best Local Similarity	49.7%;	Pred. No. 2e-70;		
Matches 152; Conservative	52;	Mismatches 99;	Gaps 3;	
				Indels 2;

[illegible]

RESULT 9  
W38217 ID W38217 standard; Protein; 320 AA.  
AC W38217;  
DT 27-Apr-1998 (first entry)  
DE *Stylosanthes humilis* Shp66 peroxidase.  
NM Peroxidase; Shp66; transgenic plant; fungus resistance

KW disease protection; Phytophthora parasitica;  
 KM Leptosphaeria maculans; Sclerotinia sclerotiorum.  
 OS Stylosanthes humilis strain Paterson.  
 FH Key Location/Qualifiers  
 FT Peptide 1. 24  
 FT /label= sig-peptide  
 PN W09741237-A1.  
 PD 06-NOV-1997.  
 PF 29-APR-1997; AU0253.  
 PF 29-APR-1996; AU-009532.  
 PA (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.  
 PA (GRAI-) GRAINS RES & DEV CORP.  
 PI Coulter KC, Kazan K, Manners JM;  
 DR WPI: 97-549739/50.  
 DR N-PSDB: T95782.  
 PT plant cell transformed with Stylosanthes humilis Shpx6 peroxidase -  
 PT useful to confer on plants resistance to fungi, e.g. Phytophthora  
 PT parasitica, Leptosphaeria maculans or Sclerotinia sclerotiorum  
 PT Claim 4; Page 15-16, 30pp; English.  
 PS This protein comprises the Shpx6 peroxidase of Stylosanthes  
 CC humilis. A claimed method of engineering a plant to fungal  
 CC resistance comprises introducing into cells of the plant a DNA  
 CC construct comprising a promoter that is constitutively operative  
 CC in the plant cell (preferably the cauliflower mosaic virus 35S  
 CC promoter) and a Shpx6 DNA sequence (see T95782), or a hybridising  
 CC sequence or fragment that encodes an enzyme with peroxidase  
 CC activity. Also claimed are plant cells harboring the DNA  
 CC construct, a plant comprising such cells, and material from such  
 CC a plant, especially seed, pollen, a stem segment or a cutting.  
 CC Resistance to fungi, especially Phytophthora parasitica,  
 CC Leptosphaeria maculans or Sclerotinia sclerotiorum, can be  
 CC conferred on cereal, legume, oilseed, sugar or fibre plants,  
 CC particularly maize, banana, peanut, field pea, sunflower, tomato,  
 CC canola, tobacco, wheat, barley, oat, potato, soybean, cotton,  
 CC carnation and sorghum.  
 SQ Sequence 320 AA;

Query Match 43.9%; Score 705.5; DB 1; Length 320;  
 Best Local Similarity 48.4%; Pred. No. 5e-64;  
 Matches 149; Conservative 51; Mismatches 93; Indels 15; Gaps 6;

[illegible]

RESULT	10
W16607	
ID	W16607 standard; Protein; 324 AA
AC	W16607;
DT	25-JUL-1997 (first entry)
DE	Soybean peroxidase [SEPA]

KM	Peroxidases; SEPpal; soybean; plant breeding; pulp; paper; bleaching;
KW	waste disposal; soil reclamation; remediation; monoclonal antibody.
OS	Glycine max cv. ResnKk.
FH	Key
FT	peptide
FT	protein
FT	peptide
FT	binding-site
FT	modified-site
FT	modified-site
FT	modified-site
FT	modified-site
FT	modified-site
FT	modified-site
FT	modified-site
FT	modified-site
FT	modified-site
PN	WO9715656-A1.
PD	01-MAY-1997.
PF	11-OCT-1996; U16354.
PR	27-OCT-1995; US-549658.
PI	(INDI-) INDIAN CROP IMPROVEMENT ASSOC.
PA	Vierling RA.
DH	WPI; 97-259011/23.
DR	N-P5DB; T66684.
PT	Isolated soybean peroxidase genes - useful for developing products for quantification and monitoring of peroxidase activity
PS	Claim 2: Page 32-34; 63pp: English.
CC	Soybean peroxidases SEPpa1, SEPpa2, SEBp1 and SEBp2 (U16607-10) are respectively encoded by cDNA clones (T66684-87) obt'd. from a soybean seed bud cDNA library. Isolation of these clones allows the prod'n. of the soybean peroxidases, free of other soybean proteins, in transformed host cells. The peroxidases can be used to raise monoclonal antibodies (Mb's). A Mb's which is preferably immunoreactive with the peroxidase and does not interfere with its enzymatic activity when bound, can be used in a non-destructive assay for peroxidase activity in plant tissues, pref. seed coat. Sequence 324 AA:

[illegible]

Db	317	KHCAFIN3	324
RESULT	11		
ID	W81493		
AC	W81493	standard; Protein; 324 AA.	
DE	29-JAN-1999	(first entry)	
DT	Soybean peroxidase SEpal polypeptide.		
KW	Soybean peroxidase; SEpal; SEpa2; SEpb1; SEpb2; recombinant;		
KW	Industry; diagnostic chemistry.		
OS	Glycine max.		
FT	Key	location/Qualifiers	
FT	Peptide	1..21	
FT	Protein	/note="signal peptide" 22..324	
FT		/note="mature protein"	
PN	US5840558-A.		
PD	24-NOV-1998.		
PF	27-OCT-1995: 671320.		
PR	27-OCT-1995: US-671320		
PA	(INDT-) INDIANA CROP IMPROVEMENT ASSOC.		
PI	Vierling RA;		
PI	WPI; 99-034041/03.		
DR	N-PSDB: V69273.		
PT	cDNA encoding soya peroxidase SEpal - useful for producing		
PT	recombinant peroxidase		
PS	Claim 1: Columns 21-24; 31pp: English.		
CC	This represents a soybean peroxidase SEpal polypeptide. The invention		
CC	provides four cDNA sequences (V69273 to V69276) encoding soybean		
CC	peroxidases SEpal, SEpa2, SEpb1 and SEpb2 (W81493 to W81496)		
CC	respectively. An expression vector containing the SEpal cDNA sequence		
CC	can be used to transform host cells for the recombinant production of		
CC	the peroxidase. The peroxidase is useful in industries and diagnostic		
CC	chemistries.		
Sequence	324 AA;		
50			

[illegible]

KM Peroxidase; SEpal; soybean; transgenic plant; immunoassay; assay;  
KM marker; pulp; paper; bleaching; bioremediation;  
OS soil decontamination.  
FH Glycine max.  
FH Key Location/Qualifiers  
FT Peptide 1..21  
FT /note= "signal peptide"  
FT Protein 22..324  
FT /note= "mature protein"  
FT Peptide 22..28  
FT /note= "QLXXFY motif, characteristic of plant peroxidases"  
FT Disulfide\_bond 32  
FT /note= "Cys residue involved in disulfide bonding"  
FT Disulfide\_bond 65  
FT /note= "Cys residue involved in disulfide bonding"  
FT Disulfide\_bond 70  
FT /note= "Cys residue involved in disulfide bonding"  
FT Disulfide\_bond 110  
FT /note= "Cys residue involved in disulfide bonding"  
FT Disulfide\_bond 116  
FT /note= "Cys residue involved in disulfide bonding"  
FT Disulfide\_bond 195  
FT /note= "Cys residue involved in disulfide bonding"  
FT Disulfide\_bond 227  
FT /note= "Cys residue involved in disulfide bonding"  
FT Disulfide\_bond 319  
FT /note= "Cys residue involved in disulfide bonding"  
FT Disulfide\_bond 319  
FT /note= "Cys residue involved in disulfide bonding"  
FT Modified\_site 77  
FT /note= "N-glycosylation"  
FT Modified\_site 90  
FT /note= "N-glycosylation"  
FT Modified\_site 149  
FT /note= "N-glycosylation"  
FT Modified\_site 163  
FT /note= "N-glycosylation"  
FT Modified\_site 204  
FT /note= "N-glycosylation"  
FT Modified\_site 235  
FT /note= "N-glycosylation"  
FT Modified\_site /note= "N-glycosylation"  
PN W0955629-A2.  
PD 10-DEC-1998.  
PE 04-JUN-1998; U11421.  
PR 04-JUN-1997; US-868577.  
PA (INDI-) INDIANA CROP IMPROVEMENT ASSOC.  
PI Vierling RA:  
DR WPI; 99-070273/06.  
DR N-PSDB; V81418 AND V81420.  
PT New soybean peroxidase genes - useful, e.g. in pulp and paper  
PS bleaching, on site waste destruction and soil remediation  
PS Example 8; Page 38-40; 78pp; English.  
CC This is the amino acid sequence of novel soybean peroxidase SEpal,  
CC as deduced from a cDNA clone (see V81420) isolated from a soybean  
CC cv. Resnick seedling cDNA library. Soybean peroxidase SEpal,  
CC SEpal2, SEpal1 and SEpal2 (see W67731..34) sequences are provided.  
CC SEpal and SEpal2 exhibit 97% amino acid identity. SEpal and SEpal2  
CC share 95% amino acid identity, and SEpal and SEpal share 47% amino  
CC acid identity. The peroxidases are useful in pulp and paper  
CC bleaching, on-site waste destruction, soil remediation, organic  
CC synthesis and diagnostic chemistries. Soybean peroxidase has  
CC advantages over chlorine bleach, being cheaper, more environmentally  
CC friendly, and producing hydroxyl ions with twice the oxidising  
CC power of chlorine ions. The plant enzyme is cheap and easy to  
CC produce. The invention also relates to immunoassays or  
CC oligonucleotide assays which utilise soybean peroxidase as marker.  
SQ Sequence 324 AA;

Query Match 37.0%; Score 594.5; DB 1; Length 324;  
Best Local Similarity 39.9%; Pred. No. 1.1e-57;  
Matches 122; Conservative 63; Mismatches 115; Indels 7; Gaps 5;

QY 2 QLTPTFDNSCPNVSNIVRDTIVNELRSDPRIASILRLHFHDFVNGCDASILLDMTTS 61

Db 22 QIQGFYANSCPKAEQIVLKFVHDIHNPISLAALILMHFHDCEVRCDCASVILNSTN 81  
QY 62 FRTEKDAFGNANSARGPEVIDRMKAAYESACPRVSCADLTITAAQOSTYLAGPSMVP 121  
Db 82 -QAEKNAPPNL-TVRGEPDIRIKSLVEACPGVASCADLTITLAARDTVAVGPGFMPVP 139  
QY 122 LGRDSDLAQFIDLANANLPAPEFTLPOLKDSFRVNGVLRSSDVALSGCHFGKNOCRFI 181  
Db 140 TGRDGVSNLTETARNNIPAPSSNFTTQTLEFANQGLD-LKDVLLSQAHTIGAHCSL 198  
QY 182 MDRLYNFSNGLPPELTNTTYLQTLRGL-CPLNGNISALVDPLRPTTFDMKRYVNL 239  
Db 199 SNRLFNFTGKGDDPSLDSEFAANILKAFKCYDLKLNKLTTKIEMPGSKTFEDLSYSHVI 258  
QY 240 EQKGLIQSDQELFSPDADTITPIVRSANSTQFFNAFVAMBRMGITPLTGTGQIR 299  
Db 259 KRRLFFESDALTLT-NSVTKAOTIILLEGSENEFFAFATSIEMGRINVKGTGEIR 316  
QY 300 LNCRVNS 307  
Db 317 KHCAPINS 324

RESULT 13  
W16608  
ID W16608 standard; Protein: 324 AA.  
AC W16608:  
DT 25-JUL-1997 (first entry)  
DE Soybean peroxidase SEpal2.  
KM waste disposal; SEpal2; soybean; plant breeding; pulp; paper; bleaching;  
OS Glycine max cv. Resnik.  
FH Key Location/Qualifiers  
FT Peptide 1..21  
FT /label= Sig\_peptide  
FT protein 22..324  
FT /label= Mat\_protein  
FT peptide 22..28  
FT /label= QLXXFY\_motif  
FT binding\_site 61..67  
FT /label= Haem\_binding\_site  
FT modified\_site 77  
FT /label= N-glycosylation\_site  
FT modified\_site 90  
FT /label= N-glycosylation\_site  
FT modified\_site 149  
FT /label= N-glycosylation\_site  
FT modified\_site 163  
FT /label= N-glycosylation\_site  
FT modified\_site 204  
FT /label= N-glycosylation\_site  
FT modified\_site 235  
FT /label= N-glycosylation\_site  
FT W09715656-A1.  
PN 04-MAY-1997.  
PE 11-OCT-1996; U16354.  
PR 27-OCT-1995; US-549658.  
PA (INDI-) INDIANA CROP IMPROVEMENT ASSOC.  
PI Vierling RA:  
DR WPI; 97-259011/23.  
DR N-PSDB; T66685.  
PT Isolated soybean peroxidase genes - useful for developing products  
PS for quantification and monitoring of peroxidase activity  
PS Claim 5; Page 36-37; 63pp; English.  
CC Soybean peroxidases SEpal, SEpal2, SEpal1 and SEpal2 (W16607-10) are  
CC respectively encoded by cDNA clones (T66684-87) obtd. from a  
CC soybean seed bud cDNA library. Isolation of these clones allows  
CC the prodn. of the soybean peroxidases, free of other soybean  
CC proteins, in transformed host cells. The peroxidases can be used  
CC to raise monoclonal antibodies (MAb). A MAb, which is preferably  
CC immunoreactive with the peroxidase and does not interfere with its  
CC enzymatic activity when bound, can be used in a non-destructive



PR 04-JUN-1997: US-868577.  
PA (INDI-) INDIANA CROP IMPROVEMENT ASSOC.  
PI Vierling RA:  
DR WPI: 99-070273/06.  
DR N-PSDB: V81421.  
PT New soybean peroxidase genes - useful, e.g. in pulp and paper  
PT bleaching, on site waste destruction and soil remediation  
PS Example 8; Page 41-43; 76pp: English.  
CC This is the amino acid sequence of novel soybean peroxidase SEp2,  
CC as deduced from a cDNA clone (see V81421) isolated from a soybean  
CC cv. Resnick seedling cDNA library. Soybean peroxidase SEp1,  
CC SEp2, SEp3 and SEp2 (see W67731-34) sequences are provided.  
CC SEp1 and SEp2 exhibit 97% amino acid identity, SEp1 and SEp2  
CC share 95% amino acid identity, and SEp1 and SEp1 share 47% amino  
CC acid identity. The peroxidases are useful in pulp and paper  
CC bleaching, on-site waste destruction, soil remediation, organic  
CC synthesis and diagnostic chemistries. Soybean peroxidase has  
CC advantages over chlorine bleach, being cheaper, more environmentally  
CC friendly, and producing hydroxyl ions with twice the oxidising  
CC power of chlorine ions. The plant enzyme is cheap and easy to  
CC produce. The invention also relates to immunoassays or  
CC oligonucleotide assays which utilise soybean peroxidase as marker.  
CC Sequence 324 AA:  
SQ

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 12:38:09 ; Search time 66.92 Seconds  
(without alignments)  
77.394 Million cell updates/sec

Title: US-09-246-451-17  
Perfect score: 1608  
Sequence: 1 MOLPTFPYDNCSPVSNIVR.....PLTGTQGIRLNCRVYNSNS 309

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 164575 seqs, 1676186 residues  
Total number of hits satisfying chosen parameters: 164575

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgnl\_7/ptodata/1/laa/5A-COMB.pep:\*  
2: /cgnl\_7/ptodata/1/laa/5B-COMB.pep:\*  
3: /cgnl\_7/ptodata/1/laa/6-COMB.pep:\*  
4: /cgnl\_7/ptodata/1/laa/PCTUS-COMB.pep:\*  
5: /cgnl\_7/ptodata/1/laa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1462	90.9	349	US-08-190-029A-10	Sequence 10, Appl
2	1462	90.9	349	US-08-462-695-10	Sequence 10, Appl
3	594.5	37.0	324	US-08-671-320-11	Sequence 11, Appl
4	594.5	37.0	324	US-08-868-577-11	Sequence 11, Appl
5	591.5	36.8	324	US-08-671-320-13	Sequence 13, Appl
6	591.5	36.8	324	US-08-868-577-13	Sequence 13, Appl
7	566	35.2	313	US-08-671-320-15	Sequence 15, Appl
8	566	35.2	313	US-08-868-577-15	Sequence 15, Appl
9	565	35.1	313	US-08-671-320-17	Sequence 17, Appl
10	565	35.1	313	US-08-868-577-17	Sequence 17, Appl
11	298.5	18.6	121	US-08-190-029A-12	Sequence 12, Appl
12	298.5	18.6	121	US-08-462-695-12	Sequence 12, Appl
13	118.5	7.4	726	US-08-313-185-49	Sequence 49, Appl
14	118.5	7.4	726	US-08-459-499-13	Sequence 13, Appl
15	116	7.2	1600	US-08-617-697-10	Sequence 10, Appl
16	111	6.9	1529	US-08-728-470-10	Sequence 10, Appl
17	110.5	6.9	294	US-08-459-499-16	Sequence 16, Appl
18	110.5	6.9	729	US-08-313-185-50	Sequence 50, Appl
19	110.5	6.9	729	US-08-459-499-14	Sequence 14, Appl
20	107	6.7	293	US-08-313-185-52	Sequence 52, Appl
21	85.5	5.3	1358	US-08-404-665-4	Sequence 4, Appl
22	85.5	5.3	1358	US-08-404-671-4	Sequence 4, Appl
23	85.5	5.3	1358	US-08-404-781-4	Sequence 4, Appl
24	84	5.2	560	US-08-756-317-8	Sequence 8, Appl
25	84	5.2	933	US-08-293-728-2	Sequence 2, Appl
26	83.5	5.2	928	US-08-474-140-11	Sequence 11, Appl
27	83.5	5.2	928	US-08-477-630-11	Sequence 11, Appl
28	83.5	5.2	928	US-08-472-293-11	Sequence 11, Appl

## ALIGNMENTS

29	83.5	5.2	928	US-08-474-545-11	Sequence 11, Appl
30	83.5	5.2	928	US-08-478-341-11	Sequence 11, Appl
31	83.5	5.2	928	US-08-996-733-11	Sequence 11, Appl
32	82	5.1	553	US-08-663-566A-13	Sequence 13, Appl
33	82	5.1	553	US-08-484-575A-14	Sequence 14, Appl
34	82	5.1	553	US-08-023-610-13	Sequence 13, Appl
35	82	5.1	553	US-08-288-065A-13	Sequence 13, Appl
36	82	5.1	553	US-08-362-240A-13	Sequence 13, Appl
37	82	5.1	553	US-08-477-459-14	Sequence 14, Appl
38	82	5.1	553	PCT-US94-01826A-14	Sequence 14, Appl
39	82	5.1	553	PCT-US94-02252A-14	Sequence 14, Appl
40	82	5.1	553	PCT-US95-10245-13	Sequence 13, Appl
41	81	5.0	638	US-08-426-125-5	Sequence 5, Appl
42	81	5.0	638	US-08-455-355-5	Sequence 5, Appl
43	80	5.0	313	US-08-373-858-2	Sequence 2, Appl
44	80	5.0	313	US-08-500-611-2	Sequence 2, Appl
45	80	5.0	313	US-08-500-694-2	Sequence 2, Appl

RESULT 1  
US-08-190-029A-10  
Sequence 10, Application US/08190029A  
Patent No. 5736363  
GENERAL INFORMATION:  
APPLICANT: EDWARDS, Richard Mark  
APPLICANT: BAWDEN, Lindsey  
TITLE OF INVENTION: IGF-II ANALOGUES  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ALLEGRETTI & MITCHELL, LTD.  
STREET: 10 S. WACKER DRIVE, SUITE 3000  
CITY: CHICAGO  
STATE: ILLINOIS  
COUNTRY: U.S.A.  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/190,029A  
FILING DATE: 28-FEB-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/G892/01389  
FILING DATE: 27-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9202401.7  
FILING DATE: 05-FEB-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9116325.3  
FILING DATE: 29-JUL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: JOHN J. McDONNELL  
REGISTRATION NUMBER: 26,949  
REFERENCE/DOCKET NUMBER: 94,062  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-715-1000  
TELEFAX: 312-715-1234  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 349 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-190-029A-10  
Query Match 90.9%; Score 1462; DB 1; Length 349;

Best Local Similarity 99.6%; Pred. No. 3e-152;  
Matches 281; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 MQLTPFYDNCSPVNSIVVDITVNELRSDPRIAASILRLHFHDCFYNGCDASILLDNTT 60
Db 1 MQLTPFYDNCSPVNSIVVDITVNELRSDPRIAASILRLHFHDCFYNGCDASILLDNTT 60
OY 61 SFPTKDAFGNANSARGFPVIDRMKAAVESACPRVSCADLLTTAAQSVTLAGGPSMRV 120
Db 61 SFPTKDAFGNANSARGFPVIDRMKAAVESACPRVSCADLLTTAAQSVTLAGGPSMRV 120
OY 121 PLGRDSLOAFDLAANALPAPEFTLPQLKDSFRNGLNSSLVALSGHTEGKNOCRF 180
Db 121 PLGRDSLOAFDLAANALPAPEFTLPQLKDSFRNGLNSSLVALSGHTEGKNOCRF 180
OY 181 IMRLVFSNTGTPDPTLNTTYLQTLRGICPLNGNLSALVDFDLRPTTFIDNKYYVNL 240
Db 181 IMRLVFSNTGTPDPTLNTTYLQTLRGICPLNGNLSALVDFDLRPTTFIDNKYYVNL 240
OY 241 QKGLIQSDQELFSSPDATDTIPLVRSFANSTQTFEFAVFEAM 282
Db 241 QKGLIQSDQELFSSPDATDTIPLVRSFANSTQTFEFAVFEAM 282
```

```
RESULT 2
US-08-462-695-10
; Sequence 10, Application US/08462695
; Patent No. 5854025
; GENERAL INFORMATION:
; APPLICANT: EDWARDS, Richard Mark
; APPLICANT: BANDEN, Lindsey
; TITLE OF INVENTION: IGF-II ANALOGUES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BANNER 6 ALLEGRETTI, LTD.
; STREET: 10 S. WACKER DRIVE, SUITE 3000
; CITY: CHICAGO
; STATE: ILLINOIS
; COUNTRY: U.S.A.
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,695
; FILING DATE: 5-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/190,029
; FILING DATE: 28-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB92/01389
; FILING DATE: 27-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9202401.7
; FILING DATE: 05-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9116325.3
; FILING DATE: 29-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: JOHN J. McDONNELL
; REGISTRATION NUMBER: 26,949
; REFERENCE/DOCKET NUMBER: 94,062-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-715-1000
; TELEFAX: 312-715-1234
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 349 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
```

; MOLECULE TYPE: protein  
US-08-462-695-10

Query Match 90.9%; Score 1462; DB 2: Length 349;  
Best Local Similarity 99.6%; Pred. No. 3e-152;  
Matches 281; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 MQLTPFYDNCSPVNSIVVDITVNELRSDPRIAASILRLHFHDCFYNGCDASILLDNTT 60
Db 1 MQLTPFYDNCSPVNSIVVDITVNELRSDPRIAASILRLHFHDCFYNGCDASILLDNTT 60
OY 61 SFPTKDAFGNANSARGFPVIDRMKAAVESACPRVSCADLLTTAAQSVTLAGGPSMRV 120
Db 61 SFPTKDAFGNANSARGFPVIDRMKAAVESACPRVSCADLLTTAAQSVTLAGGPSMRV 120
OY 121 PLGRDSLOAFDLAANALPAPEFTLPQLKDSFRNGLNSSLVALSGHTEGKNOCRF 180
Db 121 PLGRDSLOAFDLAANALPAPEFTLPQLKDSFRNGLNSSLVALSGHTEGKNOCRF 180
OY 181 IMRLVFSNTGTPDPTLNTTYLQTLRGICPLNGNLSALVDFDLRPTTFIDNKYYVNL 240
Db 181 IMRLVFSNTGTPDPTLNTTYLQTLRGICPLNGNLSALVDFDLRPTTFIDNKYYVNL 240
OY 241 QKGLIQSDQELFSSPDATDTIPLVRSFANSTQTFEFAVFEAM 282
Db 241 QKGLIQSDQELFSSPDATDTIPLVRSFANSTQTFEFAVFEAM 282
```

```
RESULT 3
US-08-671-320-11
; Sequence 11, Application US/08671320
; Patent No. 5840558
; GENERAL INFORMATION:
; APPLICANT: VIERLING JR, RICHARD A
; TITLE OF INVENTION: A SOYBEAN PEROXIDASE GENE FAMILY AND AN
; ASSAY FOR DETECTING SOYBEAN PEROXIDASE ACTIVITY
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: 1125 SO. 103RD STREET
; STREET: SUITE 330
; CITY: OMAHA
; STATE: NE
; COUNTRY: US
; ZIP: 68124-1076
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/671,320
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JONDLE, ROBERT J.
; REGISTRATION NUMBER: 33,915
; REFERENCE/DOCKET NUMBER: 1227-001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 402-398-9000
; TELEFAX: 402-398-9005
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 324 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-671-320-11
```

Query Match 37.0%; Score 594.5; DB 2: Length 324;  
Best Local Similarity 89.9%; Pred. No. 4.5e-57;  
Matches 123; Conservative 63; Mismatches 115; Indels 7; Gaps 5;

[illegible]

RESULT 4  
 US-08-868-577-11  
 ; Sequence 11, Application US/08868577  
 ; Patent No. 5866695  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Vierling Jr., Richard A  
 ; TITLE OF INVENTION: A SOYBEAN PEROXIDASE GENE FAMILY AND AN  
 ; TITLE OF INVENTION: ASSAY FOR DETECTING SOYBEAN PEROXIDASE ACTIVITY  
 ; NUMBER OF SEQUENCES: 19  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Rothwell, Figg, Ernst & Kurz  
 ; STREET: 555 13th Street NW, Suite 701 East  
 ; City: Washington  
 ; STATE: DC  
 ; COUNTRY: USA  
 ; ZIP: 20004  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Wordperfect 6.1  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/868,577  
 ; FILING DATE: 04-JUN-1997  
 ; CLASSIFICATION: 536  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Jondle, Robert J.  
 ; REGISTRATION NUMBER: 33,915  
 ; REFERENCE/DOCKET NUMBER: NJ227-003  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 402-333-1550  
 ; TELEFAX: 402-333-1510  
 ; INFORMATION FOR SEQ ID NO: 11:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 324 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-868-577-11

Query Match	37.0%	Score 594.5;	DB 2;	Length 324;
Best Local Similarity	39.9%;	Pred. No.	4.5e-57;	
Matches 123;	Conservative 63;	Mismatches 115;	Indels 7;	Gaps 5;
QY	2 QLTFFFNOSCPCNVSNVTRDTIVNLRSDPRAAISILRHFDHFCVNGCDASILLDNNTTS	61		
	: : :   :::::               ::::   :			

Db 22 QIAGGFANSCPKKEQIVLKFVHDHINNAASLAALAIRMHFHCFVGCASVILNSTN 81

QY 62 FTRKKDAPGANSARGPEVIDRMKAAYEASCRTVSGADILTTIAAQOSVTLACGSMWRP 121

Db 82 -QAKKNMAPPNLTVRGDFIDRIKSLVEAECPGVASCADILTTIAARTIYATGGPEMKVP 139

QY 122 LGRDSDLOAFDLNANLPAPEFTLPOLKDSFRVNGINRSSDVLVSGHTFGKNOCRF 181

Db 140 TGRDDGVASNLTEARNNIPAPSSNFTLTQLTFANQGLD-LKDLVLLSGAHTTIGAHCSL 198

QY 182 MDRILXNSNGLCPPTPLNTIYTLQTLRCL-CPLNGNLSALVDPLRLPTTFDMKYYNLE 239

Db 199 SNRLPNTGTGKDQDPSLDSELAANLAKFKCTDLKLKNTFYKIEPDGSRKTFDLSYSHVI 258

QY 240 EOKGLISDDELFPSSPATPTIPLVRFANSFSTOFFENAFPEADRMKNITPLTGTGQIR 299

Db 259 KRKGLEPSDALLLT-NSVTKAQITQLLEGSVENEFAEFATSIEMKGRINWKTGTGEIR 316

QY 300 LNCRAVNS 307

Db 317 KHCAPLNS 324

```

1      RESULT      5
2      US-08-671-320-13
3      : Sequence 13: Application US/08671320
4      : Patent No. 5840558
5      : GENERAL INFORMATION:
6      : APPLICANT: VIERLING JR, RICHARD A
7      : TITLE OF INVENTION: A SOYBEAN PEROXIDASE GENE FAMILY AND AN
8      : TITLE OF INVENTION: ASSAY FOR DETECTING SOYBEAN PEROXIDASE ACTIVITY
9      : NUMBER OF SEQUENCES: 17
10     : CORRESPONDENCE ADDRESS:
11     : ADDRESSEE: 1125 SO. 103RD STREET
12     : STREET: SUITE 330
13     : CITY: OMAHA
14     : STATE: NE
15     : COUNTRY: US
16     : ZIP: 68124-1076
17     : COMPUTER READABLE FORM:
18     : MEDIUM TYPE: Floppy disk
19     : COMPUTER: IBM PC compatible
20     : OPERATING SYSTEM: PC-DOS/MS-DOS
21     : SOFTWARE: PatentIn Release #1.0, Version #1.30
22     : CURRENT APPLICATION DATA:
23     : APPLICATION NUMBER: US/08/671,320
24     : FILING DATE:
25     : CLASSIFICATION: 435
26     : ATTORNEY/AGENT INFORMATION:
27     : NAME: JONDLE, ROBERT J.
28     : REGISTRATION NUMBER: 33,915
29     : REFERENCE/DOCKET NUMBER: 1227-001
30     : TELECOMMUNICATION INFORMATION:
31     : TELEPHONE: 402-398-9000
32     : TELEFAX: 402-398-9005
33     : INFORMATION FOR SEQ ID NO: 13:
34     : SEQUENCE CHARACTERISTICS:
35     : LENGTH: 324 amino acids
36     : TYPE: amino acid
37     : TOPOLOGY: linear
38     : MOLECULE TYPE: protein
39     : US-08-671-320-13

```

[illegible]

Db 82 -QAKNAPPNL-IVRGDFIDRIKSLVEAECPGVASCADILITLSAROTIVATGGPEMKVP 139  
122 LGRDLSQAFELDLANANLPAPEFTLPOLKDSFRNVGLNRSSDLVALSGHTFGKNOCRFI 181  
140 TGRDGYISNLTEARDONI PAPSSNFTTLQTLFANOGLD-LKDLVLLSGAHTIGIAHCSSL 198  
QY 182 MDRLYNSNGLPDPPTLNTTYTLQTLRGL--CPLNGNL SALVDFDLRPTTFDNKYVNL 239  
199 SNRLFNFTGKGDDPSIDSEYAA NLKAFKCTDLNKLNTTTEMDPGSRKTFDLSYSHVI 258  
QY 240 EOKGLISDOELFSSPATDTIPLVRSFANSSTQTFENAFV EADMGMNITPLTGOQOIR 299  
Db 259 KRGLEFSDAALLT--NSVTKAQIIEELGSEVENFEAFATSMKGRINVKGTGEIR 316  
QY 300 LNCRVNS 307  
Db 317 KHCAFLNS 324

RESULT 6  
US-08-868-577-13  
; Sequence 13, Application US/08868577  
; Patent No. 5866695  
; GENERAL INFORMATION:  
; APPLICANT: VIERLING JR., RICHARD A  
; TITLE OF INVENTION: A SOYBEAN PEROXIDASE GENE FAMILY AND AN  
; TITLE OF INVENTION: ASSAY FOR DETECTING SOYBEAN PEROXIDASE ACTIVITY  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz  
; STREET: 555 13th Street NW, Suite 701 East  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: WordPerfect 6.1  
; CURRENT APPLICATION DATA:  
; FILING DATE: 04-JUN-1997  
; APPLICATION NUMBER: US/08/868,577  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jondle, Robert J.  
; REGISTRATION NUMBER: 33,915  
; REFERENCE/DOCKET NUMBER: N1227-003  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 402-333-1550  
; TELEFAX: 402-333-1510  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 324 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-868-577-13

Query Match 36.8%; Score 591.5; DB 2; Length 324;  
Best Local Similarity 39.9%; Pred. No. 9.5e-57;  
Matches 123; Conservative 62; Mismatches 116; Indels 7; Gaps 5;

QY 122 LGRDLSQAFELDLANANLPAPEFTLPOLKDSFRNVGLNRSSDLVALSGHTFGKNOCRFI 181  
140 TGRDGYISNLTEARDONI PAPSSNFTTLQTLFANOGLD-LKDLVLLSGAHTIGIAHCSSL 198  
QY 182 MDRLYNSNGLPDPPTLNTTYTLQTLRGL--CPLNGNL SALVDFDLRPTTFDNKYVNL 239  
199 SNRLFNFTGKGDDPSIDSEYAA NLKAFKCTDLNKLNTTTEMDPGSRKTFDLSYSHVI 258  
QY 240 EOKGLISDOELFSSPATDTIPLVRSFANSSTQTFENAFV EADMGMNITPLTGOQOIR 299  
Db 259 KRGLEFSDAALLT--NSVTKAQIIEELGSEVENFEAFATSMKGRINVKGTGEIR 316  
QY 300 LNCRVNS 307  
Db 317 KHCAFLNS 324

RESULT 7  
US-08-671-320-15  
; Sequence 15, Application US/08671320  
; Patent No. 5840558  
; GENERAL INFORMATION:  
; APPLICANT: VIERLING JR, RICHARD A  
; TITLE OF INVENTION: A SOYBEAN PEROXIDASE GENE FAMILY AND AN  
; TITLE OF INVENTION: ASSAY FOR DETECTING SOYBEAN PEROXIDASE ACTIVITY  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: 1125 SO. 103RD STREET  
; STREET: SUITE 330  
; CITY: OMAHA  
; STATE: NE  
; COUNTRY: US  
; ZIP: 68124-1076  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE:  
; APPLICATION NUMBER: US/08/671,320  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: JONDLE, ROBERT J.  
; REGISTRATION NUMBER: 33,915  
; REFERENCE/DOCKET NUMBER: 1227-001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 402-398-9000  
; TELEFAX: 402-398-9005  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 313 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-671-320-15

Query Match 35.2%; Score 566; DB 2; Length 313;  
Best Local Similarity 44.4%; Pred. No. 5.7e-54;  
Matches 127; Conservative 44; Mismatches 107; Indels 8; Gaps 6;

QY 3 LPTFTDSCPNVSNIVRTIVNELRSDPRIASILRLHFHDFVNGCDASILDNTTSP 62  
Db 23 LSLNTYAKTCPNVEFYIAKAAVKATARDKTVPAIILRMHFHDFVNGCDASVLLNSKNN 82  
QY 63 RTEKDAFGNANSARGPVIDRMAKAVESACPTVSCADILITTAOQSVTLGAGPSRVPL 122  
Db 83 KAEKDEDPNV-SLHAFYVIAAKKALEASCPGVASCADILALAAVAVFLSGPTMDVVK 141  
QY 123 GRRDLSQAFELDLANANLPAPEFTLPOLKDSFRNVGLNRSSDLVALSGHTFGKNOCRFI 182  
Db 142 GRRDGRSKAS-ETROLPAPEFTNLSQLRSFSQGLS-GBDLVALSGHTLGFSSFK 199

```

0Y 183 DRVFNSTGLPDLPLNTTYIQTGRGCLP-NGULSLAVPEDLRTPIPNKKYVHLEQ 241
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 200 NRHHFNNTHVDVSLNSFPAKILISTICPLKNAOKNKGTSMDST-TTPNTTYRLLLOO 258
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
0Y 242 KGLIOSDQELFSPDADDTLPVRSFANSQTQFFNAFVEAMDRCGN 287
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 259 KGLFSSDDVLLDNP--DTKKLNTYRTSKKAFLEAFKASIRNRS 301
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

1      RESULT 8
2      US-08-868-577-15
3      ; Sequence 15, Application US/08868577
4      ; Patent No. 5866695
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Vierling Jr., Richard A
7      ; TITLE OF INVENTION: A SOYBEAN PEROXIDASE GENE FAMILY AND AN
8      ; TITLE OF INVENTION: ASSAY FOR DETECTING SOYBEAN PEROXIDASE ACTIVITY
9      ; NUMBER OF SEQUENCES: 19
10     ; CORRESPONDENCE ADDRESS:
11     ; ADDRESSEE: Rothwell, Fidy, Ernst & Kurz
12     ; STREET: 555 13th Street NW, Suite 701 East
13     ; CITY: Washington
14     ; STATE: DC
15     ; COUNTRY: USA
16     ; ZIP: 20004
17     ; COMPUTER READABLE FORM:
18     ; MEDIUM TYPE: Floppy disk
19     ; COMPUTER: IBM PC compatible
20     ; OPERATING SYSTEM: PC-DOS/MS-DOS
21     ; SOFTWARE: WordPerfect 6.1
22     ; CURRENT APPLICATION DATA:
23     ; APPLICATION NUMBER: US/08/868,577
24     ; FILING DATE: 04-JUN-1997
25     ; CLASSIFICATION: 536
26     ; ATTORNEY/AGENT INFORMATION:
27     ; NAME: Jondle, Robert J.
28     ; REGISTRATION NUMBER: 33,915
29     ; REFERENCE/DOCKET NUMBER: N1227-003
30     ; TELECOMMUNICATION INFORMATION:
31     ; TELEPHONE: 402-333-1550
32     ; TELEFAX: 402-333-1510
33     ; INFORMATION FOR SEQ ID NO: 15:
34     ; SEQUENCE CHARACTERISTICS:
35     ; LENGTH: 313 amino acids
36     ; TYPE: amino acid
37     ; TOPOLOGY: linear
38     ; MOLECULE TYPE: protein
39     ; US-08-868-577-15

```

[illegible]

Db 259 KGLFSSDQVLLDNP---DTKNLVTKFATSKKAFYEAFKSMIRMS 301

```

1  RESULT 9
2  US-08-671-320-17
3  ; Sequence 17, Application US/08671320
4  ; Patent No. 5840558
5  ;
6  ; GENERAL INFORMATION:
7  ;
8  ; APPLICANT: VIERLING JR, RICHARD A
9  ;
10 ; TITLE OF INVENTION: A SOYBEAN PEROXIDASE GENE FAMILY AND AN
11 ; TITLE OF INVENTION: ASSAY FOR DETECTING SOYBEAN PEROXIDASE ACTIVITY
12 ; NUMBER OF SEQUENCES: 17
13 ;
14 ; CORRESPONDENCE ADDRESS:
15 ; ADDRESSEE: 1125 SO. 103RD STREET
16 ; STREET: SUITE 330
17 ; CITY: OMAHA
18 ;
19 ; STATE: NE
20 ;
21 ; COUNTRY: US
22 ;
23 ; ZIP: 68124-1076
24 ;
25 ; COMPUTER READABLE FORM:
26 ;
27 ; MEDIUM TYPE: floppy disk
28 ;
29 ; COMPUTER: IBM PC compatible
30 ;
31 ; OPERATING SYSTEM: PC-DOS/MS-DOS
32 ;
33 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
34 ;
35 ; CURRENT APPLICATION DATA:
36 ;
37 ; APPLICATION NUMBER: US/08/671,320
38 ;
39 ; FILING DATE:
40 ;
41 ; CLASSIFICATION: 435
42 ;
43 ; ATTORNEY/AGENT INFORMATION:
44 ;
45 ; NAME: JONDELE, ROBERT J.
46 ;
47 ; REGISTRATION NUMBER: 33,915
48 ;
49 ; REFERENCE/DOCKET NUMBER: 1227-001
50 ;
51 ; TELECOMMUNICATION INFORMATION:
52 ;
53 ; TELEPHONE: 402-398-9000
54 ;
55 ; TELEFAX: 402-398-9005
56 ;
57 ; INFORMATION FOR SEQ ID NO: 17:
58 ;
59 ; SEQUENCE CHARACTERISTICS:
60 ;
61 ; LENGTH: 313 amino acids
62 ;
63 ; TYPE: amino acid
64 ;
65 ; TOPOLOGY: linear
66 ;
67 ; MOLECULE TYPE: protein
68 ;
69 ;
70 ;
71 ;
72 ;
73 ;
74 ;
75 ;
76 ;
77 ;
78 ;
79 ;
80 ;
81 ;
82 ;
83 ;
84 ;
85 ;
86 ;
87 ;
88 ;
89 ;
90 ;
91 ;
92 ;
93 ;
94 ;
95 ;
96 ;
97 ;
98 ;
99 ;
100 ;
101 ;
102 ;
103 ;
104 ;
105 ;
106 ;
107 ;
108 ;
109 ;
110 ;
111 ;
112 ;
113 ;
114 ;
115 ;
116 ;
117 ;
118 ;
119 ;
120 ;
121 ;
122 ;
123 ;
124 ;
125 ;
126 ;
127 ;
128 ;
129 ;
130 ;
131 ;
132 ;
133 ;
134 ;
135 ;
136 ;
137 ;
138 ;
139 ;
140 ;
141 ;
142 ;
143 ;
144 ;
145 ;
146 ;
147 ;
148 ;
149 ;
150 ;
151 ;
152 ;
153 ;
154 ;
155 ;
156 ;
157 ;
158 ;
159 ;
160 ;
161 ;
162 ;
163 ;
164 ;
165 ;
166 ;
167 ;
168 ;
169 ;
170 ;
171 ;
172 ;
173 ;
174 ;
175 ;
176 ;
177 ;
178 ;
179 ;
180 ;
181 ;
182 ;
183 ;
184 ;
185 ;
186 ;
187 ;
188 ;
189 ;
190 ;
191 ;
192 ;
193 ;
194 ;
195 ;
196 ;
197 ;
198 ;
199 ;
200 ;
201 ;
202 ;
203 ;
204 ;
205 ;
206 ;
207 ;
208 ;
209 ;
210 ;
211 ;
212 ;
213 ;
214 ;
215 ;
216 ;
217 ;
218 ;
219 ;
220 ;
221 ;
222 ;
223 ;
224 ;
225 ;
226 ;
227 ;
228 ;
229 ;
230 ;
231 ;
232 ;
233 ;
234 ;
235 ;
236 ;
237 ;
238 ;
239 ;
240 ;
241 ;
242 ;
243 ;
244 ;
245 ;
246 ;
247 ;
248 ;
249 ;
250 ;
251 ;
252 ;
253 ;
254 ;
255 ;
256 ;
257 ;
258 ;
259 ;
260 ;
261 ;
262 ;
263 ;
264 ;
265 ;
266 ;
267 ;
268 ;
269 ;
270 ;
271 ;
272 ;
273 ;
274 ;
275 ;
276 ;
277 ;
278 ;
279 ;
280 ;
281 ;
282 ;
283 ;
284 ;
285 ;
286 ;
287 ;
288 ;
289 ;
290 ;
291 ;
292 ;
293 ;
294 ;
295 ;
296 ;
297 ;
298 ;
299 ;
300 ;
301 ;
302 ;
303 ;
304 ;
305 ;
306 ;
307 ;
308 ;
309 ;
310 ;
311 ;
312 ;
313 ;
314 ;
315 ;
316 ;
317 ;
318 ;
319 ;
320 ;
321 ;
322 ;
323 ;
324 ;
325 ;
326 ;
327 ;
328 ;
329 ;
330 ;
331 ;
332 ;
333 ;
334 ;
335 ;
336 ;
337 ;
338 ;
339 ;
340 ;
341 ;
342 ;
343 ;
344 ;
345 ;
346 ;
347 ;
348 ;
349 ;
350 ;
351 ;
352 ;
353 ;
354 ;
355 ;
356 ;
357 ;
358 ;
359 ;
360 ;
361 ;
362 ;
363 ;
364 ;
365 ;
366 ;
367 ;
368 ;
369 ;
370 ;
371 ;
372 ;
373 ;
374 ;
375 ;
376 ;
377 ;
378 ;
379 ;
380 ;
381 ;
382 ;
383 ;
384 ;
385 ;
386 ;
387 ;
388 ;
389 ;
390 ;
391 ;
392 ;
393 ;
394 ;
395 ;
396 ;
397 ;
398 ;
399 ;
400 ;
401 ;
402 ;
403 ;
404 ;
405 ;
406 ;
407 ;
408 ;
409 ;
410 ;
411 ;
412 ;
413 ;
414 ;
415 ;
416 ;
417 ;
418 ;
419 ;
420 ;
421 ;
422 ;
423 ;
424 ;
425 ;
426 ;
427 ;
428 ;
429 ;
430 ;
431 ;
432 ;
433 ;
434 ;
435 ;
436 ;
437 ;
438 ;
439 ;
440 ;
441 ;
442 ;
443 ;
444 ;
445 ;
446 ;
447 ;
448 ;
449 ;
450 ;
451 ;
452 ;
453 ;
454 ;
455 ;
456 ;
457 ;
458 ;
459 ;
460 ;
461 ;
462 ;
463 ;
464 ;
465 ;
466 ;
467 ;
468 ;
469 ;
470 ;
471 ;
472 ;
473 ;
474 ;
475 ;
476 ;
477 ;
478 ;
479 ;
480 ;
481 ;
482 ;
483 ;
484 ;
485 ;
486 ;
487 ;
488 ;
489 ;
490 ;
491 ;
492 ;
493 ;
494 ;
495 ;
496 ;
497 ;
498 ;
499 ;
500 ;
501 ;
502 ;
503 ;
504 ;
505 ;
506 ;
507 ;
508 ;
509 ;
510 ;
511 ;
512 ;
513 ;
514 ;
515 ;
516 ;
517 ;
518 ;
519 ;
520 ;
521 ;
522 ;
523 ;
524 ;
525 ;
526 ;
527 ;
528 ;
529 ;
530 ;
531 ;
532 ;
533 ;
534 ;
535 ;
536 ;
537 ;
538 ;
539 ;
540 ;
541 ;
542 ;
543 ;
544 ;
545 ;
546 ;
547 ;
548 ;
549 ;
550 ;
551 ;
552 ;
553 ;
554 ;
555 ;
556 ;
557 ;
558 ;
559 ;
560 ;
561 ;
562 ;
563 ;
564 ;
565 ;
566 ;
567 ;
568 ;
569 ;
570 ;
571 ;
572 ;
573 ;
574 ;
575 ;
576 ;
577 ;
578 ;
579 ;
580 ;
581 ;
582 ;
583 ;
584 ;
585 ;
586 ;
587 ;
588 ;
589 ;
590 ;
591 ;
592 ;
593 ;
594 ;
595 ;
596 ;
597 ;
598 ;
599 ;
600 ;
601 ;
602 ;
603 ;
604 ;
605 ;
606 ;
607 ;
608 ;
609 ;
610 ;
611 ;
612 ;
613 ;
614 ;
615 ;
616 ;
617 ;
618 ;
619 ;
620 ;
621 ;
622 ;
623 ;
624 ;
625 ;
626 ;
627 ;
628 ;
629 ;
630 ;
631 ;
632 ;
633 ;
634 ;
635 ;
636 ;
637 ;
638 ;
639 ;
640 ;
641 ;
642 ;
643 ;
644 ;
645 ;
646 ;
647 ;
648 ;
649 ;
650 ;
651 ;
652 ;
653 ;
654 ;
655 ;
656 ;
657 ;
658 ;
659 ;
660 ;
661 ;
662 ;
663 ;
664 ;
665 ;
666 ;
667 ;
668 ;
669 ;
670 ;
671 ;
672 ;
673 ;
674 ;
675 ;
676 ;
677 ;
678 ;
679 ;
680 ;
681 ;
682 ;
683 ;
684 ;
685 ;
686 ;
687 ;
688 ;
689 ;
690 ;
691 ;
692 ;
693 ;
694 ;
695 ;
696 ;
697 ;
698 ;
699 ;
700 ;
701 ;
702 ;
703 ;
704 ;
705 ;
706 ;
707 ;
708 ;
709 ;
710 ;
711 ;
712 ;
713 ;
714 ;
715 ;
716 ;
717 ;
718 ;
719 ;
720 ;
721 ;
722 ;
723 ;
724 ;
725 ;
726 ;
727 ;
728 ;
729 ;
730 ;
731 ;
732 ;
733 ;
734 ;
735 ;
736 ;
737 ;
738 ;
739 ;
740 ;
741 ;
742 ;
743 ;
744 ;
745 ;
746 ;
747 ;
748 ;
749 ;
750 ;
751 ;
752 ;
753 ;
754 ;
755 ;
756 ;
757 ;
758 ;
759 ;
760 ;
761 ;
762 ;
763 ;
764 ;
765 ;
766 ;

```

```

Query Match Similarity 35.1%; Score 565; DB 2; Length 313;
Best Local Similarity 43.6%; Pred. NO. 7.3e-54;
Matches 125; Conservative 47; Mismatches 107; Indels 8; Gaps 6;

QY 3 LNPTEYDNCSPVNSIVRDTIVNELRSPDIRIASILRLHFHDCFPVNGCDASILDNTTSF 62
   | : : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 23 LSLNATYKTCPCPVECLVAKAVKADATAROKDTYPAALLRMHFHDCFPVNGCDASVLNSKGN 82

QY 63 RTEQAFGNANSARGCPVYIDRMKAAVESACPTVSCADLLITIAAQSYYTLACGSPSWRVL 122
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 83 KAEKGGPPRV - SLHAFYVIDAKKAALEASCSPGVSCADITALLAARDVAFLSGGPTWDEPK 141

QY 123 GARDSLQARLDLANANLPAFPFTLQJLQKDSFPNNCGINRSDDVALVSLGGTHPTKKNOCRFM 182
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 142 GKKDCTSKAS - ETRQLPAPFTNLSQLRFSFORGIS - GEDLVASGGTHPTGFSHCSSPK 199

QY 183 DRLVYNSNTGCLPDPPTLNTTYLQTLKGLCP - GGNLSALVDFCLRTPTIDNRKYVNLREQ 241
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 200 NRIHNPNAIHHDDPSLNSFATKLISICLKKQAKNAGSMDPSI - TTTDNTYYIKLILQO 258

QY 242 KGLIOSDOELFSSPDATDTIPLVRSFANSSTQTFENAFVEMDRKNI 288
   | | | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 259 KGLFSSDQVLLDNP - - - DTKNLVAKAFATSKAFYDAFAKSMIKMSSI 302

RESULT 10
US-08-868-577-17
; Sequence 17, Application US/08868577
; Patent No. 5866695

```

```

; GENERAL INFORMATION:
; APPLICANT: VIERLING JR., Richard A
; TITLE OF INVENTION: A SOYBEAN PEROXIDASE GENE FAMILY AND AN
; TITLE OF INVENTION: ASSAY FOR DETECTING SOYBEAN PEROXIDASE ACTIVITY
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz
; STREET: 555 13th Street NW, Suite 701 East
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/868,577
; FILING DATE: 04-JUN-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Jondle, Robert J.
; REGISTRATION NUMBER: 33,915
; REFERENCE/DOCKET NUMBER: N1227-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 402-333-1550
; TELEFAX: 402-333-1510
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 313 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-868-577-17

Query Match 35.1%; Score 565; DB 2; Length 313;
Best Local Similarity 43.6%; Pred. No. 7.3e-54;
Matches 125; Conservative 47; Mismatches 107; Indels 8; Gaps 6;

QY 3 LTPFVNSCPNYSNIVRDTIVNELRSDPRIASILRLHFDGCVNGCDASILLDNTTSF 62
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 23 LSNLYSKTPEDECTVAKAKDPTAPALRLMHFHDGCVNGCDASVLLNSGNS 82
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 63 RTEKDAFGNANSARGFEVIDRMKAVESACPRVSCADLLTTIAAQSQVTLAGGSMVPL 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 83 KAEDGPVNT-SLHAFVVIDAAKKALEASCPGVVSCADLLALAAADVFLSGGTWDEPK 141
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 123 GRDSDLAFLDLANANLPAPFTLLPOLKDSFRVNGLNRRSSDLVALSGGHTFGKNOCREFIM 182
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 142 GRKDGRTSKAS-ETROLPAPTFNLSQLRQSFQSGRLS-GEIDLVALSGGHTLGFSSCSFK 199
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 183 DRLYNFSNGLPDPPLNTNTYIQTLRGLCP-LNGLSLALVDFDLRTPIEDNKYYVNLREQ 241
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 200 NRIHNENRATDEDESLPSPATKILISICPLKNAKNAKGTMSDST-TTFNTFYRLILQQ 258
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 242 KGLIOSDQELFSSPDATDTIPLVRSFANSTOTFFNAFVEADMRGNI 288
   ||| ||| | : | : | : | : | : | : | : | : | : | : | : | : |
DB 259 KGLFSSDQVLLDNP---DTKNLVAKAFATSKAFYDAFAKSMIMMSI 302
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
; US-08-190-029A-12
; Sequence 12, Application US/08190029A
; Patent No. 5736363
; GENERAL INFORMATION:
; APPLICANT: EDWARDS, Richard Mark
; APPLICANT: BAWDEN, Lindsey
; TITLE OF INVENTION: IGF-II ANALOGUES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: ALLEGRETTI & WITCOFF, LTD.
; STREET: 10 S. WACKER DRIVE, SUITE 3000
; CITY: CHICAGO
; STATE: ILLINOIS
; COUNTRY: U.S.A.
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/190,029A
; FILING DATE: 28-FEB-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB92/01389
; FILING DATE: 27-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9202401.7
; FILING DATE: 05-FEB-1992
; APPLICATION NUMBER: GB 9116325.3
; FILING DATE: 29-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: JOHN J. McDONNELL
; REGISTRATION NUMBER: 26,949
; REFERENCE/DOCKET NUMBER: 94,062
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-715-1000
; TELEFAX: 312-715-1234
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 121 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-190-029A-12

Query Match 18.6%; Score 298.5; DB 1; Length 121;
Best Local Similarity 56.2%; Pred. No. 2.9e-25;
Matches 68; Conservative 6; Mismatches 26; Indels 21; Gaps 3;

QY 1 MQLTFEYDSCPNYSNIVRDTIVNELRSDPRIASILRLHFDGCVNGCDASILLDNTT 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1 MQLTFEYDSCPNYSNIVRDTIVNELRSDPRIASILRLHFDGCVNGCDASIM----- 54
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 61 SEFTKDAFQN-----ANSARGF-----PVIDRMKAVESACPRVSCADLLTTA 105
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 55 AYRPSFTLCGGLVDTLQFVCGDGRGFYSRPAVSRSRSGIVEECFCRCDLALLETYC 114
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 106 A 106
   |
DB 115 A 115
   |

RESULT 12
; US-08-462-695-12
; Sequence 12, Application US/08462695
; Patent No. 5854025
; GENERAL INFORMATION:
; APPLICANT: EDWARDS, Richard Mark
; APPLICANT: BAWDEN, Lindsey
; TITLE OF INVENTION: IGF-II ANALOGUES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: BANNER & ALLEGRETTI, LTD.
; STREET: 10 S. WACKER DRIVE, SUITE 3000
; CITY: CHICAGO
; STATE: ILLINOIS
; COUNTRY: U.S.A.
; ZIP: 60606
; COMPUTER READABLE FORM:
```

```

; STREET: 10 S. WACKER DRIVE, SUITE 3000
; CITY: CHICAGO
; STATE: ILLINOIS
; COUNTRY: U.S.A.
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/190,029A
; FILING DATE: 28-FEB-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB92/01389
; FILING DATE: 27-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9202401.7
; FILING DATE: 05-FEB-1992
; APPLICATION NUMBER: GB 9116325.3
; FILING DATE: 29-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: JOHN J. McDONNELL
; REGISTRATION NUMBER: 26,949
; REFERENCE/DOCKET NUMBER: 94,062
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-715-1000
; TELEFAX: 312-715-1234
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 121 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-190-029A-12

Query Match 18.6%; Score 298.5; DB 1; Length 121;
Best Local Similarity 56.2%; Pred. No. 2.9e-25;
Matches 68; Conservative 6; Mismatches 26; Indels 21; Gaps 3;

QY 1 MQLTFEYDSCPNYSNIVRDTIVNELRSDPRIASILRLHFDGCVNGCDASILLDNTT 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1 MQLTFEYDSCPNYSNIVRDTIVNELRSDPRIASILRLHFDGCVNGCDASIM----- 54
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 61 SEFTKDAFQN-----ANSARGF-----PVIDRMKAVESACPRVSCADLLTTA 105
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 55 AYRPSFTLCGGLVDTLQFVCGDGRGFYSRPAVSRSRSGIVEECFCRCDLALLETYC 114
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 106 A 106
   |
DB 115 A 115
   |

RESULT 12
; US-08-462-695-12
; Sequence 12, Application US/08462695
; Patent No. 5854025
; GENERAL INFORMATION:
; APPLICANT: EDWARDS, Richard Mark
; APPLICANT: BAWDEN, Lindsey
; TITLE OF INVENTION: IGF-II ANALOGUES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: BANNER & ALLEGRETTI, LTD.
; STREET: 10 S. WACKER DRIVE, SUITE 3000
; CITY: CHICAGO
; STATE: ILLINOIS
; COUNTRY: U.S.A.
; ZIP: 60606
; COMPUTER READABLE FORM:
```

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/462,695  
CLASSIFICATION: 435  
FILING DATE: 5-JUN-1995  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 08/190,029  
FILING DATE: 28-FEB-1994  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB92/01389  
FILING DATE: 27-JUL-1992  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: GB 9202401.7  
FILING DATE: 05-FEB-1992  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: GB 9116325.3  
FILING DATE: 29-JUL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: JOHN J. McDONNELL  
REGISTRATION NUMBER: 26,949  
REFERENCE/DOCKET NUMBER: 94,062-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-715-1000  
TELEFAX: 312-715-1234  
INFORMATION FOR SEQ. ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 121 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-462-695-12

Query Match 18.6%; Score 298.5; DB 2; Length 121;  
Best Local Similarity 56.2%; Pred. No. 2.9e-25;  
Matches 68; Conservative 6; Mismatches 26; Indels 21; Gaps 3;  
QY 1 MQLPTFYDMSCPVNSIVVDITVNELRSDPRIASLRLHFHDCFPVNGCDASILLDNTT 60  
DB 1 MQLPTFYDMSCPVNSIVVDITVNELRSDPRIASLRLHFHDCFPVNGCDASM----- 54  
QY 61 SPFRKDAFGN-----ANSARGF-----PYIDRMKAAYESACPRIVSCADLLTIA 105  
DB 55 AYRPSEILCGELVDTLQFYVCGDRGFYFRPASRVSRRSRSGIYECCFRSCDLALLETC 114  
QY 106 A 106  
DB 115 A 115  
RESULT 13  
US-08-313-185-49  
Sequence 49, Application US/08313185  
Patent No. 5851763  
GENERAL INFORMATION:  
APPLICANT: Heym, Beate  
APPLICANT: Cole, Stewart  
APPLICANT: Young, Douglas  
APPLICANT: Zhang, Ying  
APPLICANT: Honore, Nadine  
APPLICANT: Teienti, Amalio  
TITLE OF INVENTION: Rapid Detection of Antibiotic Resistance  
TITLE OF INVENTION: In Mycobacterium Tuberculosis  
NUMBER OF SEQUENCES: 66  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Parabow, Garrett &  
ADDRESSEE: Dunner  
STREET: 1300 I Street, N.W.  
CITY: Washington

STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/313,185  
FILING DATE: 12-OCT-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 02356.0068-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4000  
TELEFAX: (202) 408-4400  
INFORMATION FOR SEQ. ID NO: 49:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 726 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-313-185-49

Query Match 7.4%; Score 118.5; DB 2; Length 726;  
Best Local Similarity 23.1%; Pred. No. 0.00029;  
Matches 61; Conservative 24; Mismatches 88; Indels 91; Gaps 9;  
QY 96 VSCADLTITAAQSVTLAGPSRVRPLGRDLSQAFID-----LANAN 138  
DB 157 ISWADLFILAGNVALENGSRTEFGAGRDVMPDLDVWNGDEKAWLTHRHPEALAKAP 216  
QY 139 LPA-----PFTTLPOLKDSFNVGLNRSDDLVALSGCHTFEKNOCRF 180  
DB 217 LGATEMGLIYVPEGPHPHSGEPLSAAMAIIRTFEGNMGNDEYVALIAGGHTLAKTH--- 273  
QY 181 IMRLVNSMTGLPDPPLNTTYTLTGCLPLNGN-----ISALVDFDLRPTIPDNK 233  
DB 274 ---GAPPTSNVG--PDPAAPLIEBGL-GWASTYSGVGADAITSGLEWVMTPTQMSNY 328  
QY 234 YVYVL-----EKGGLIOSDQELFSSPDATITP----- 262  
DB 329 FFENLFFYEWVQTRSPAGAIO-----FEAVDAPPEIIPDPDPKRRKPTMLVTDLTFRD 383  
QY 263 ----LVRSPANSOTFFNAPVEA 281  
DB 384 PEFKISRFLNDPQAFNEAFARA 407  
RESULT 14  
US-08-459-499-13  
Sequence 13, Application US/08459499  
Patent No. 5871912  
GENERAL INFORMATION:  
APPLICANT: Heym, Beate  
APPLICANT: Cole, Stewart T.  
APPLICANT: Young, Douglas B.  
APPLICANT: Zhang, Ying  
TITLE OF INVENTION: Nucleic Acid Probes, Sequences, and Methods  
TITLE OF INVENTION: for Detecting Mycobacterium Tuberculosis Resistant to Isoni  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Parabow, Garrett &  
ADDRESSEE: Dunner  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: DC





---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 12:36:44 ; Search time 73.09 Seconds

(without alignments)  
261.628 Million cell updates/sec

Title: US-09-246-451-17

Perfect score: 1608

Sequence: 1 MQLPFTFDNSCPVNSNIVR.....PLTGNQGIRLNCRVNSNS 309

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 178050 seqs, 61884766 residues

Total number of hits satisfying chosen parameters: 178050

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : PIR-64:\*\*

1: p1r1:\*\*  
2: p1r2:\*\*  
3: p1r3:\*\*  
4: p1r4:\*\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1598	99.4	308	2 PC7025	peroxidase (EC 1.1
2	1598	99.4	353	1 OPRHC	peroxidase (EC 1.1
3	1509	93.8	353	2 T46118	peroxidase - Arabi
4	1500	93.3	353	2 S37495	peroxidase (EC 1.1
5	1475	91.7	354	2 JU0457	peroxidase (EC 1.1
6	1473	91.6	332	2 S00627	peroxidase (EC 1.1
7	1469	91.4	351	2 S00626	peroxidase (EC 1.1
8	1271	79.0	347	2 JH0149	peroxidase (EC 1.1
9	1125.5	70.0	349	2 JH0150	peroxidase (EC 1.1
10	1108.5	68.9	349	2 T02506	peroxidase (EC 1.1
11	1079.5	67.1	349	2 T02507	peroxidase (EC 1.1
12	1065.5	66.3	349	2 JU0458	peroxidase (EC 1.1
13	1046	65.0	343	2 T09566	peroxidase (EC 1.1
14	1044	64.9	343	2 T09565	peroxidase (EC 1.1
15	1002	62.3	343	2 S60055	peroxidase (EC 1.1
16	975.5	60.7	347	2 S60054	peroxidase (EC 1.1
17	921	57.3	352	2 T05723	peroxidase (EC 1.1
18	918	57.1	353	2 T09665	peroxidase (EC 1.1
19	910	56.6	351	2 J04779	peroxidase (EC 1.1
20	906	56.3	355	2 J04780	peroxidase (EC 1.1
21	894.5	55.6	358	2 J04781	peroxidase (EC 1.1
22	787	48.9	318	2 J02217	peroxidase (EC 1.1
23	786.5	48.9	324	2 A39889	peroxidase (EC 1.1
24	780.5	48.5	322	2 T03686	peroxidase (EC 1.1
25	779	48.4	325	2 S32768	peroxidase (EC 1.1
26	768	47.8	327	2 S14268	peroxidase (EC 1.1
27	758.5	47.2	336	2 T45730	peroxidase-like pr
28	752.5	46.8	316	2 S61405	peroxidase (EC 1.1
29	746.5	46.4	314	2 S61406	peroxidase (EC 1.1

30	742.5	46.2	316	2 A38265	peroxidase (EC 1.1
31	742	46.1	315	2 T09165	probable peroxidase
32	738	45.9	312	2 S13375	peroxidase (EC 1.1
33	737.5	45.9	292	2 S11870	peroxidase (EC 1.1
34	729.5	45.4	312	2 S13325	peroxidase (EC 1.1
35	729.5	45.4	319	2 S61408	peroxidase (EC 1.1
36	727	45.2	305	2 T08121	peroxidase (EC 1.1
37	724	45.0	329	2 T10444	peroxidase (EC 1.1
38	715	44.5	315	2 T06172	peroxidase (EC 1.1
39	714	44.4	314	2 T03929	peroxidase (EC 1.1
40	712.5	44.3	335	2 T03912	peroxidase (EC 1.1
41	711.5	44.2	335	2 T04333	peroxidase (EC 1.1
42	710	44.2	317	2 S22087	peroxidase (EC 1.1
43	708.5	44.1	283	2 T06778	peroxidase (EC 1.1
44	707	44.0	315	2 T06164	peroxidase (EC 1.1
45	702	43.7	322	2 B56555	peroxidase (EC 1.1

ALIGNMENTS

Query Match	99.4%	Score 1598	DB 2	Length 308
Best Local Similarity	99.7%	Pred. No. 3.7e-128		
Matches 307	Conservative 1	Mismatches 0	Indels 0	Gaps 0
QY	2 QLPFTFDNSCPVNSNIVRPIYVELRSDPRIASIRLHDFVNGCASILLDDTTS 61			
DB	1 QLPFTFDNSCPVNSNIVRPIYVELRSDPRIASIRLHDFVNGCASILLDDTTS 60			
QY	62 FTEKDAFGNANSARCPVIDRMKAAYESACPTVSCADLLTTIAAQSVTLAGSPSRVP 121			
DB	61 FTEKDAFGNANSARCPVIDRMKAAYESACPTVSCADLLTTIAAQSVTLAGSPSRVP 120			
QY	122 LGRDSTQAFDLANANLPAFFETLPLQKDSFRVNGLRSSDVALSGHTFGKNOCRFI 181			
DB	121 LGRDSTQAFDLANANLPAFFETLPLQKDSFRVNGLRSSDVALSGHTFGKNOCRFI 180			
QY	182 MDRLYNSNTGLPPTNTTYLQTLRGICPLNGMLSAVDFDLTPPIFPNKYYVNLDEQ 241			
DB	181 MDRLYNSNTGLPPTNTTYLQTLRGICPLNGMLSAVDFDLTPPIFPNKYYVNLDEQ 240			
QY	242 KGLIQSQELFSSPDATDTPIVRSFANSOTQTFNATVEAMDNRGNTPLTGTGQIRLN 301			
DB	241 KGLIQSQELFSSPDATDTPIVRSFANSOTQTFNATVEAMDNRGNTPLTGTGQIRLN 300			
QY	302 CRVYNSNS 309			
DB	301 CRVYNSNS 308			
RESULT	2			
OPRHC	peroxidase (EC 1.11.1.7) CIA precursor - horseradish			
C:Species:	Armoracia rusticana (horseradish)			
C:Date:	24-Apr-1984 #sequence_revision 12-Apr-1996 #text_change 03-Mar-2000			

C:Accession: S00625; S32972; A00502  
R:Fujiyama, K.; Takemura, H.; Shibayama, S.; Kobayashi, K.; Choi, J.K.; Shimmyo, A.; Tak  
Eur. J. Biochem. 173, 681-687, 1988  
A:Title: Structure of the horseradish peroxidase isozyme C genes.  
A:Reference number: S00625; MUID:8825087  
A:Accession: S00625  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-353 <FUJ>  
R:Wellinder, K.G.  
Eur. J. Biochem. 96, 483-502, 1979  
A:Title: Amino acid sequence studies of horseradish peroxidase. Amino and carboxyl termi  
radish peroxidase C. S32972; MUID:79236311  
A:Reference number: S32972; MUID:79236311  
A:Accession: S32972  
A:Molecule type: protein  
A:Residues: 31-338 <WEL>  
R:Wellinder, K.G.  
FEBS Lett. 72, 19-23, 1976  
A:Title: Covalent structure of the glycoprotein horseradish peroxidase (EC 1.11.1.7).  
A:Reference number: A00502; MUID:77068850  
A:Accession: A00502  
A:Molecule type: protein  
A:Residues: 31-338 <WEL2>  
C:Genetics:  
A:Gene: prxC1  
A:Superfamily: peroxidase  
C:Keywords: chromoprotein; glycoprotein; heme; iron; metalloprotein; oxidoreductase; py  
F.1-30/Domain: signal sequence #status predicted <SIG>  
F.31-338/Product: peroxidase C1 #status experimental <MAT>  
F.31/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experim  
F.41-121,74-79,127-331,207-239/Disulfide bonds: #status experimental  
F.43,87,188,216,228,244,285,298/Binding site: carbohydrate (Asn) (covalent) #status exp  
F.68/Active site: Arg #status predicted  
F.72,200/Binding site: heme iron (His) (axial ligands) #status predicted

Query Match 99.4%; Score 1598; DB 1; Length 353;  
Best Local Similarity 99.7%; Pred. No. 4,5e-128;  
Matches 307; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 QLTPTFYDNCSPVNSIVRTIVNELSDPRIASILRLHEDCVNGCDASILDNTTS 61  
DB 31 QLTPTFYDNCSPVNSIVRTIVNELSDPRIASILRLHEDCVNGCDASILDNTTS 90  
QY 62 FRTKFAFGNANSARGPVIDRMAKAAVESACPRTVSCADLLTIAQOQSYTLAGGSPWRVP 121  
DB 91 FRTKFAFGNANSARGPVIDRMAKAAVESACPRTVSCADLLTIAQOQSYTLAGGSPWRVP 150  
QY 122 LGRDRLQAFLLDANANLPAFFTLPLQKDSFRVNGLNRSSDLVALSGGHTFGKNQCRPT 181  
DB 151 LGRDRLQAFLLDANANLPAFFTLPLQKDSFRVNGLNRSSDLVALSGGHTFGKNQCRPT 210  
QY 182 MDRLVNFSGTGLPDPILNTTYYLQTLRGLCLPLNGNLALVDLFDLRTPTTFDNKYVNLLEEQ 241  
DB 211 MDRLVNFSGTGLPDPILNTTYYLQTLRGLCLPLNGNLALVDLFDLRTPTTFDNKYVNLLEEQ 270  
QY 242 KGLIGSDQELFSSPDATDTLPIVNSFANSQTFFNAPFAMDGMKNTPLTGTOGIRLN 301  
DB 271 KGLIGSDQELFSSPDATDTLPIVNSFANSQTFFNAPFAMDGMKNTPLTGTOGIRLN 330  
QY 302 CRVYVNSNS 309  
DB 331 CRVYVNSNS 338

RESULT 3  
T46118  
peroxidase - Arabidopsis thaliana  
N:Alternate names: protein T2173.40  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 04-Feb-2000 #sequence\_rev1sion 04-Feb-2000 #text\_change 18-Feb-2000  
C:Accession: T46118

R.Rieger, M.; Gabel, C.; Mueller-Auer, S.; Schaefer, M.; Zipp, M.; Mewes, H.W.; Mayer submitted to the Protein Sequence Database, November 1999

A:Reference number: Z23023

A:Accession: T46118

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-353 <RIE>

A:Cross-references: EMBL:AL132967

A:Experimental source: cultivar Columbia; BAC clone T2J13

C:Genetics:

A:Map position: 3

A:Introns: 77/3; 141/3; 198/1

A:Note: T2J13.40

C:Superfamily: peroxidase

```

Query Match          93.8%; Score 1509; DB 2; Length 353;
Best Local Similarity 92.9%; Pred. No. 1,66-120;
Matches 286; Conservative 15; Mismatches 7; Indels 0; Gaps 0;

QY 2 QLTPEFYDNCGVNSINIVRDTIVNELRSDPRIAASILRLHFHDFVNCGDASILNDNTS 61
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 31 QLTPEFYRSCGVNIVIRETIVNELRSDPRIAASILRLHFHDFVNCGDASILNDNTS 90
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 62 FTTEKDAFGNANSARGFVIDRMKAAYSACPRVYSCADLLTTAAQSVTLAGBSWRVP 121
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 91 FTTEKDAFGNANSARGFVIDRMKAAYERACPRVYSCADMLTTAAQSVTLAGBSWRVP 150
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 122 IGRRSLSAFLDILANANIPAPFTLTPOLKDSFRNGLRSSDVALSGHFPGNOCRFI 181
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 151 IGRRSLSAFLDILANANIPAPFTLTPOLKASFRNGLRPSDVALSGHFPGNOCRFI 210
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 182 MDRLYNFSNTGLPDEPTLNTTYLQTLRGCLPNGNLISALVDEDLRTPTTFEDMKRYVNLREQ 241
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 211 LDRLYNFSNTGLPDEPTLNTTYLQTLRGCLPNGNSALVDEDLRTPTTFEDMKRYVNLKER 270
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 242 KGLIOSDDELSSPATTTIPLVRSFANSTQTFRAAFPEANDRMGNITPLTGTGQOIRLN 301
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 271 KGLIOSDDELSSPATDTIPLVRAVADGTQTFENAFVEMNRGNITPTGTGQOIRLN 330
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 302 CRVYVNS 309
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 331 CRVYVNS 338

RESULT 4
S37495
peroxidase (EC 1.11.1.7) Cb - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 09-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 04-Mar-2000
C:Accession: S37495
R:Infapruk, C.; Takano, M.; Shimmyo, A.
submitted to the EMBL Data Library, April 1993
A:Description: Nucleotide sequence of a new cDNA for peroxidase from Arabidopsis thal
A:Reference number: S37495
A:Accession: S37495
A:Molecule type: mRNA
A:Residues: 1-553 <INT>
A:Cross-references: EMBL:X71794; NID:g405610; PIDN:CAA50677.1; PID:g405611
C:Genetics:
A:Gene: prpxcb
C:Superfamily: peroxidase
C:Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase
F:41-121/Dissulfide bonds: #status predicted
F:68/Active site: Arg #status predicted
F:72/200/Blinding site: heme iron (His) (axial ligands) #status predicted
F:74-79/Dissulfide bonds: #status predicted
F:127-131/Dissulfide bonds: #status predicted
F:207-239/Dissulfide bonds: #status predicted

Query Match          93.3%; Score 1500; DB 2; Length 353;
Best Local Similarity 92.2%; Pred. No 9,36-120;

```

```
Matches 284; Conservative 15; Mismatches 9; Indels 0; Gaps 0;
QY 2 QLPPTFYDNCSPVNSNIVRDTIVNELRSDPRIASILRLHFHOCFVNGCASILLDMTTS 61
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 31 QLPPTFYDRCSPVNTIVRETIVNELRSDPRIASILRLHFHOCFVNGCASILLDMTTS 90
QY 62 FRTEKDAFGNANSARGPVIDRMKAAVESACPRVSCADLLTTAAQOSVTLAGPSMRVP 121
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 91 FRTEKDRFGNANSARGPVIDRMKAVERACPRVSCADMLTTAAQOSVTLAGPSMRVP 150
QY 122 LGRDLSIQAFLLDANANLPAFFTLPOLKDSFRVNGLNRSDDLVALSGHTFGKNOCRFI 181
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 151 LGRDLSIQAFLELANANLPAFFTLPOLKASFRVNGLDPSDLVALSGHTFGKNOCQFI 210
QY 182 MDRLYNSNTGLPDPPLNTNTYLTQTLRGCLPLNGNLSALVDFDLRTPTIFDNKYYVNLKEQ 241
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 211 LDRFYNSNTGLPDPPLNTNTYLTQTLRGCLPLNGNRSALVDFDLRTPTIFVFNKYYVNLKER 270
QY 242 KGLIQSDQELFSSPDADTITPLVRSFANSQTQTFEFAVEAMDRMGNTTPLTGOGQIRLN 301
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 271 KGLIQSDQELFSSPNATDTIPLVRAVADGTQTFEFAVEAMNRMGNITPTTGOGQIRLN 330
QY 302 CRVYNSNS 309
   ||||| |||||
Db 331 CRVYNSNS 338

RESULT 5
peroxidase (EC 1.11.1.7) C - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 04-Mar-2000
C:Accession: J00457; T46119
R:Rintprink, C.; Higashimura, N.; Yamamoto, K.; Okada, N.; Shimmyo, A.; Takano, M.
Gene 98, 237-241, 1991
A:Title: Nucleotide sequences of two genomic DNAs encoding peroxidase of Arabidopsis tha
A:Reference number: J00457; M0ID:91200671
A:Accession: J00457
A:Molecule type: DNA
A:Residues: 1-354 <INT>
A:Cross-references: GB:M58380; NID:q16826; PIDN:AAA2849.1; PID:q16827
R:Rieger, M.; Gabel, C.; Mueller-Auer, S.; Schaefer, M.; Zipp, M.; Mewes, H.W.; Mayer, K
Submitted to the Protein Sequence Database, November 1999
A:Reference number: 223023
A:Accession: T46119
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-354 <RUE>
A:Cross-references: EMBL:AL132967
A:Experimental source: cultivar Columbia; BAC clone T2J13
C:Genetics:
A:Gene: prxCa; T2J13.50
A:Map position: 3
A:Introns: 78/3; 142/3; 199/1
C:Superfamily: peroxidase
C:Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase
F:42-122/Disulfide bonds: #status predicted
F:59/Active site: Arg #status predicted
F:73,201/Binding site: heme iron (His) (axial ligands) #status predicted
F:75-80/Disulfide bonds: #status predicted
F:128-332/Disulfide bonds: #status predicted
F:208-240/Disulfide bonds: #status predicted

Query Match 91.7%; Score 1475; DB 2; Length 354;
Best Local Similarity 90.9%; Pred. No. 1.2e-117;
Matches 280; Conservative 13; Mismatches 15; Indels 0; Gaps 0;
QY 2 QLPPTFYDNCSPVNSNIVRDTIVNELRSDPRIASILRLHFHOCFVNGCASILLDMTTS 61
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 32 QLPPTFYDTCSPVTNTVNRDTIVNELRSDPRIASILRLHFHOCFVNGCASILLDMTTS 91
QY 62 FRTEKDAFGNANSARGPVIDRMKAVERACPRVSCADLLTTAAQOSVTLAGPSMRVP 121
```

```
Db 92 FRTEKDALGNANSARGPVIDRMKAVERACPRVSCADMLTTAAQOSVTLAGPSMRVP 151
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 122 LGRDLSIQAFLLDANANLPAFFTLPOLKDSFRVNGLNRSDDLVALSGHTFGKNOCRFI 181
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 152 LGRDLSIQAFLELANANLPAFFTLPOLKANFKVNGLDPSDLVALSGHTFGKNOCRFI 211
QY 182 MDRLYNSNTGLPDPPLNTNTYLTQTLRGCLPLNGNLSALVDFDLRTPTIFDNKYYVNLKEQ 241
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 212 MDRLYNSNTGLPDPPLNTNTYLTQTLRGCCPRNGQSVLVDFDLRTPTIFVFNKYYVNLKEQ 271
QY 242 KGLIQSDQELFSSPDADTITPLVRSFANSQTQTFEFAVEAMDRMGNTTPLTGOGQIRLN 301
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 272 KGLIQSDQELFSSPNATDTIPLVRAVADGTQTFEFAVEAMNRMGNITPTTGOGQIRLN 331
QY 302 CRVYNSNS 309
   ||||| |||||
Db 332 CRVYNSNS 339

RESULT 6
peroxidase (EC 1.11.1.7) C1C precursor - horseradish (fragment)
S00627
N:Alternate names: peroxidase C3
C:Species: Armoracia rusticana (horseradish)
C:Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 04-Mar-2000
C:Accession: S00627
R:Fujiyama, K.; Takemura, H.; Shibayama, S.; Kobayashi, K.; Choi, J.K.; Shimmyo, A.;
Eur. J. Biochem. 173, 681-687, 1988
A:Title: Structure of the horseradish peroxidase isozyme C genes.
A:Reference number: S00625; M0ID:88225087
A:Accession: S00627
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-332 <FUJ>
A:Cross-references: GB:M60729; NID:q168244; PIDN:AAA33379.1; PID:q168245
C:Genetics:
A:Gene: prxC1C
C:Superfamily: peroxidase
C:Keywords: chromoprotein; glycoprotein; heme; iron; metalloprotein; oxidoreductase
F:1-9/Domain: signal sequence (fragment) #status predicted <SIG>
F:10-317/Product: peroxidase C1C #status predicted <MAT>
F:20-100/Disulfide bonds: #status predicted
F:47/Active site: Arg #status predicted
F:51,179/Binding site: heme iron (His) (axial ligands) #status predicted
F:53-58/Disulfide bonds: #status predicted
F:106-310/Disulfide bonds: #status predicted
F:186-218/Disulfide bonds: #status predicted

Query Match 91.6%; Score 1473; DB 2; Length 332;
Best Local Similarity 90.9%; Pred. No. 1.7e-117;
Matches 280; Conservative 13; Mismatches 15; Indels 0; Gaps 0;
QY 2 QLPPTFYDNCSPVNSNIVRDTIVNELRSDPRIASILRLHFHOCFVNGCASILLDMTTS 61
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10 QLPPTFYDNCSPVNSNIVRODTIVNELRSDPRIASILRLHFHOCFVNGCASILLDMTTS 69
QY 62 FRTEKDAFGNANSARGPVIDRMKAVERACPRVSCADLLTTAAQOSVTLAGPSMRVP 121
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 70 FRTEKDAFGNANSARGPVIDRKAVERACPRVSCADVLTTAAQOSVTLAGPSMRVP 129
QY 122 LGRDLSIQAFLLDANANLPAFFTLPOLKDSFRVNGLNRSDDLVALSGHTFGKNOCRFI 181
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 130 LGRDLSIQAFLELANANLPAFFTLPELKAFAVNGLNRSDDLVALSGHTFGKNOCRFI 189
QY 182 MDRLYNSNTGLPDPPLNTNTYLTQTLRGCLPLNGNLSALVDFDLRTPTIFDNKYYVNLKEQ 241
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 190 MDRLYNSNTGLPDPPLNTNTYLTQTLRQCCPRNGQSVLVDFDLRTPTIFVFNKYYVNLKEQ 249
QY 242 KGLIQSDQELFSSPDADTITPLVRSFANSQTQTFEFAVEAMDRMGNTTPLTGOGQIRLN 301
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 250 KGLIQSDQELFSSPNATDTIPLVRSYADGTQTFEFAVEAMNRMGNITPTTGOGQIRLN 309
```

QY 302 CRVYNSNS 309  
|||||||  
Db 310 CRVYNSNS 317

## RESULT 7

S00626  
peroxidase (EC 1.11.1.7) C1B precursor - horseradish  
N:Alternate names: peroxidase C2  
C:Species: Armoracia rusticana (horseradish)  
C:Date: 30-Sep-1989 #sequence\_revision 30-Sep-1989 #text\_change 04-Mar-2000  
C:Accession: S00626  
R:Fujiyama, K.; Takemura, H.; Shibayama, S.; Kobayashi, K.; Choi, J.K.; Shimmyo, A.; Takemura, H.; Takemura, H.; Shimmyo, A.; Okada, H.; Takano, M.  
A:Title: Structure of the horseradish peroxidase isozyme C genes.  
A:Reference number: S00625; MUID:88225087  
A:Accession: S00626  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-351 <FUJ>  
A:Cross-references: GB:M37157; NID:g168242; PIDN:AAA33378.1; PID:g168243  
C:Genetics:  
A:Gene: prxC1B  
C:Superfamily: peroxidase  
C:Keywords: chromoprotein; glycoprotein; heme; iron; metalloprotein; oxidoreductase  
F:1-28/Domain: signal sequence #status predicted <SIG>  
F:29-336/Product: peroxidase C1B #status predicted <MAT>  
F:39-119/Disulfide bonds: #status predicted  
F:66/Active site: Arg #status predicted  
F:70,198/Binding site: heme iron (His) (axial ligands) #status predicted  
F:125-329/Disulfide bonds: #status predicted  
F:205-237/Disulfide bonds: #status predicted

Query Match 91.4%; Score 1469; DB 2; Length 351;  
Best local Similarity 90.6%; Pred. No. 3,9e-117;

Matches 279; Conservative 12; Mismatches 17; Indels 0; Gaps 0;

QY 2 QLPFTYDNSCPVNSINVRDTYVNELRSDPRIASILRLHFHDCFYNGCDASILLDNNTS 61  
||||||| ||||||| :||||||| ||||||||| ||||||||| ||||||||| |||||||||  
Db 29 QLPFTYDSCPVNSINVRDTYVNELRSDPRIASILRLHFHDCFYNGCDASILLDNNTS 88  
QY 62 FRTKDAFGNANSARGEPYIDRKAAVESACPTVSCADLLTTAAQSVTLAGGPMRWVP 121  
||||||| ||||||| :||||||| ||||||||| ||||||||| ||||||||| |||||||||  
Db 89 FLEKDALGNANSARGEPYIDRKAAVESACPTVSCADLLTTAAQSVTLAGGPMRWVP 148  
QY 122 LGRDLSLQAFLLDANANLPAFPFTLPOLKDSFRNNGLNSSDVLVALSGHTFGKNOCRT 181  
||||||| ||||||| :||||||| ||||||||| ||||||||| ||||||||| |||||||||  
Db 149 LGRDLSLQAFLLDANANLPAFPFTLPOLKDSFRNNGLNSSDVLVALSGHTFGKNOCRT 208  
QY 182 MDRLYFNSNTGLPDPPTLNTTYLQTLGCLPLNGNLSALVDFDRLPTTFIDNKYYVNLREQ 241  
||||||| ||||||| :||||||| ||||||||| ||||||||| ||||||||| |||||||||  
Db 209 MDRLYFNSNTGLPDPPTLNTTYLQTLGCLPLNGNLSALVDFDRLPTTFIDNKYYVNLREQ 268  
QY 242 KGLIQSDQELFSSPDADTITPLVRSFANSQTFFNMFVEMDMNGNITPLTGQGIIRLN 301  
||||||| ||||||| :||||||| ||||||||| ||||||||| ||||||||| |||||||||  
Db 269 KGLIQSDQELFSSPDADTITPLVRSFANSQTFFNMFVEMDMNGNITPLTGQGIIRLN 328  
QY 302 CRVYNSNS 309  
|||||||  
Db 329 CRVYNSNS 336

## RESULT 8

JH0149  
peroxidase (EC 1.11.1.7) C2 precursor - horseradish  
C:Species: Armoracia rusticana (horseradish)  
C:Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 04-Mar-2000  
C:Accession: JH0149  
R:Fujiyama, K.; Takemura, H.; Shimmyo, A.; Okada, H.; Takano, M.  
Gene 89, 163-169, 1990

A:Title: Genomic DNA structure of two new horseradish-peroxidase-encoding genes.

A:Reference number: JH0149; MUID:90323613

A:Accession: JH0149

A:Molecule type: DNA

A:Residues: 1-347 <FUJ>

A:Cross-references: GB:D90115; NID:g217931; PIDN:BAAL4143.1; PID:d1014846; PID:g21793

C:Genetics:

A:Gene: prxC2

A:Introns: 71/3; 135/3; 192/1

C:Superfamily: peroxidase

C:Keywords: chromoprotein; glycoprotein; heme; iron; metalloprotein; oxidoreductase

F:1-24/Domain: signal sequence #status predicted <SIG>

F:25-347/Product: peroxidase C2 #status predicted <MAT>

F:35-115,68-73,121-325,201-233/Disulfide bonds: #status predicted

F:62/Active site: Arg #status predicted

F:81,210,238,310/Binding site: heme iron (His) (axial ligands) #status predicted

F:86,217,243/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 79.0%; Score 1271; DB 2; Length 347;  
Best local Similarity 79.4%; Pred. No. 2,4e-100;

Matches 243; Conservative 24; Mismatches 39; Indels 0; Gaps 0;

QY 2 QLPFTYDNSCPVNSINVRDTYVNELRSDPRIASILRLHFHDCFYNGCDASILLDNNTS 61  
||:|||| :|| :|| ||||||||| ||||||||| ||||||||| |||||||||  
Db 25 QLPSPFYDKTCPQVFPDIATNTITKALRSDPRIASILRLHFHDCFYNGCDASILLDNNTS 84  
QY 62 FRTKDAFGNANSARGEPYIDRKAAVESACPTVSCADLLTTAAQSVTLAGGPMRWVP 121  
||||||| ||||||| :||||||| ||||||||| ||||||||| ||||||||| |||||||||  
Db 85 FRTKDAFGNANSARGEPYIDRKAAVESACPTVSCADLLTTAAQSVTLAGGPMRWVP 144  
QY 122 LGRDLSLQAFLLDANANLPAFPFTLPOLKDSFRNNGLNSSDVLVALSGHTFGKNOCRT 181  
||||||| :||||| ||| ||| ||||||| ||||||||| ||||||||| |||||||||  
Db 145 LGRDLSLQAFLLDANANLPAFPFTLPOLKDSFRNNGLNSSDVLVALSGHTFGKNOCRT 204  
QY 182 MDRLYFNSNTGLPDPPTLNTTYLQTLGCLPLNGNLSALVDFDRLPTTFIDNKYYVNLREQ 241  
||||||| :||||| ||| ||| ||||||| ||||||||| ||||||||| |||||||||  
Db 205 MDRLYFNSNTGLPDPPTLNTTYLQTLGCLPLNGNLSALVDFDRLPTTFIDNKYYVNLREQ 264  
QY 242 KGLIQSDQELFSSPDADTITPLVRSFANSQTFFNMFVEMDMNGNITPLTGQGIIRLN 301  
||||||| ||||||| :||||||| ||||||||| ||||||||| ||||||||| |||||||||  
Db 265 KGLIQSDQELFSSPDADTITPLVRSFANSQTFFNMFVEMDMNGNITPLTGQGIIRLN 324  
QY 302 CRVYNS 307  
|||||||  
Db 325 CRVYNS 330

## RESULT 9

JH0150  
peroxidase (EC 1.11.1.7) C3 precursor - horseradish  
C:Species: Armoracia rusticana (horseradish)  
C:Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 04-Mar-2000  
C:Accession: JH0150  
R:Fujiyama, K.; Takemura, H.; Shimmyo, A.; Okada, H.; Takano, M.  
Gene 89, 163-169, 1990  
A:Title: Genomic DNA structure of two new horseradish-peroxidase-encoding genes.  
A:Reference number: JH0149; MUID:90323613  
A:Accession: JH0150  
A:Molecule type: DNA  
A:Residues: 1-349 <FUJ>  
A:Cross-references: GB:D90116; NID:g217933; PIDN:BAAL4144.1; PID:d1014847; PID:g21793  
C:Genetics:  
A:Gene: prxC3  
A:Introns: 76/3; 140/3; 197/1  
C:Superfamily: peroxidase  
C:Keywords: chromoprotein; glycoprotein; heme; iron; metalloprotein; oxidoreductase  
F:1-29/Domain: signal sequence #status predicted <SIG>  
F:30-349/Product: peroxidase C3 #status predicted <MAT>  
F:40-120,73-78,126-329,206-238/Disulfide bonds: #status predicted  
F:67/Active site: Arg #status predicted  
F:71,199/Binding site: heme iron (His) (axial ligands) #status predicted  
F:86,217,243/Binding site: carbohydrate (asn) (covalent) #status predicted



DB	329	CRVYN	333	
RESULT	12			
J00458				
peroxidase (EC 1.11.1.7) E - Arabidopsis thaliana				
C:Species: Arabidopsis thaliana (mouse-ear cress)				
C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 04-Mar-2000				
C:Accession: J00458				
R:Inapark, C.; Higashimura, N.; Yamamoto, K.; Okada, N.; Shimyo, A.; Takano, M.				
Gene 98, 237-241, 1991				
A:Title: Nucleotide sequences of two genomic DNA encoding peroxidase of Arabidopsis thaliana				
A:Reference number: J00457; MUID:91200671				
A:Accession: J00458				
A:Molecule type: DNA				
A:Residues: 1-349 <INT>				
A:Cross-references: GB:M58381; MID:g166806; PIDN:AAA32842.1; PID:g166807				
C:Genetics:				
A:Gene: prx6a				
A:Introns: 76/3: 140/3: 197/1				
C:Superfamily: peroxidase				
C:Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase				
F:40-120/Disulfide bonds: #status predicted				
F:67/Active site: Arg #status predicted				
F:71,199/Binding site: heme iron (His) (axial ligands) #status predicted				
F:73-78/Disulfide bonds: #status predicted				
F:126-329/Disulfide bonds: #status predicted				
F:206-238/Disulfide bonds: #status predicted				
Query Match	66.3%	Score 1065.5;	DB 2;	Length 349;
Best Local Similarity	67.5%;	Pred. No. 6.7e-83;		
Matches 206;	Conservative 37;	Mismatches 61;	Indels 1;	Gaps 1;
QY	2	QLPTPTDMSCPVNSNIVRPTVIELMSDPRIASILFLHFDCCVNGCDASILLDNITS	61	
DB	30	QLRPDYFGCPCPEVDIFGNIIVDELITDPRIASLRLHFDCCVNGCDASILLDNITS	89	
QY	62	FTEKDAFGNANSARGPVIDRMKAAVESACPRVSCADLLTTIAQOQSVTLAGGSPMRP	121	
DB	90	FTEKDAAPRANARGNNIVADRMKVALERACPCRVSCADLLTTIAQOISVLISGCPMMPVP	149	
QY	122	IGRDSLQAFLDLANANLPAPFTLPOLKDSFRVNGVLRSSDVALSGGHFGKNCRPT	181	
DB	150	KRRDSVEAFALANALPSPFENLTQLKTAFAVGLKRTSDVALSGGHFGRAQCFV	209	
QY	182	MRLVNFSTMTGLPDPITNTYTYLQTLRGICPLNGNSALVDDLPRTPTTFDNKTYVNL	241	
DB	210	TTRLVNFNGTNSDPSLXPRTYVELRLCLCPONGSTVLYVNDVTPDAFDSQYTNLRNG	269	
QY	242	KGLIQSDQELFSSPDATDTITPLVRSFANSQTGFENAFVAMD RMGNITPLTGTGQIRLN	301	
DB	270	KGLIQSDQELFSSPGA-DTITPLVNGYSSDMSVFFRAFDAMIRMGNLRPLTGTGGRIGN	328	
QY	302	CRVYN 306		
DB	329	CRVYN 333		
RESULT	13			
T09566				
peroxidase (EC 1.11.1.7) - black poplar				
C:Species: Populus nigra (black poplar)				
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999				
C:Accession: T09566				
R:Sakuma, Y.; Azuma, T.; Kato, Y.; Kojima, Y.; Miura, K.				
submitted to the EMBL Data Library, January 1996				
A:Description: Poplar peroxidase gene responding to cytokinin treatment.				
A:Reference number: Z16737				
A:Accession: T09566				
A:Molecule type: DNA				
A:Status: preliminary; translated from GB/EMBL/DBJ				
A:Molecule type: DNA				

[illegible]





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 13:04:55 ; Search time 27.02 Seconds

(Without alignments)  
354.391 Million cell updates/sec

Title: US-09-246-451-17

Perfect score: 1608

Sequence: 1 MQLPTFYDNCSPVNSNIVR.....PLTGTGQIIRLNCRVVNSNS 309

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 85661 seqs, 30989116 residues

Total number of hits satisfying chosen parameters: 85661

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_38.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1598	99.4	353	PERA_ARMRU	P00433 armoracia r
2	1475	91.7	354	PERC_ARATH	P24101 arabidopsis r
3	1473	91.6	332	PERC_ARMRU	P15233 armoracia r
4	1469	91.4	351	PERB_ARMRU	P15232 armoracia r
5	1271	79.0	347	PER2_ARMRU	P17179 armoracia r
6	1125.5	70.0	349	PER2_ARMRU	P17160 armoracia r
7	1065.5	66.3	349	PERE_ARMRU	P24102 arabidopsis r
8	875	54.4	305	PER1_ARMRU	P80679 armoracia r
9	786.5	48.9	324	PERX_ARMRU	P11965 nicotiana t
10	745.5	46.4	316	PER1_ARMRU	P19135 cucumis sal
11	737.5	45.9	292	PER1_ARMRU	P05885 trilicium ae
12	729.5	45.4	312	PER1_ARMRU	P37835 oryza sativ
13	714	44.4	314	PER2_ARMRU	P22196 hordeum vul
14	707	44.0	315	PER1_ARMRU	P00434 brassica ra
15	702	43.7	322	PERX_ARMRU	P00434 brassica ra
16	696.5	43.3	296	PERX_ARMRU	P22196 hordeum vul
17	618.5	38.5	330	PER1_ARMRU	P37834 arachis hyp
18	587.5	36.8	326	PER1_ARMRU	P37834 arachis hyp
19	463	28.5	364	PER1_ARMRU	P15003 lycopersico
20	462	28.7	351	PERX_ARMRU	P12437 solanum tub
21	440	27.4	363	PER1_ARMRU	P15004 lycopersico
22	297	18.5	158	PERX_ARMRU	P16147 lupinus pol
23	243.5	15.1	170	PER2_ARMRU	P001548 hordeum vul
24	203	12.6	80	PER1_ARMRU	P15984 trilicium ae
25	164	10.2	249	APX1_ARMRU	P48534 plasm saliv
26	156	9.7	249	APX1_ARMRU	P05431 arabidopsis
27	118.5	7.4	726	CATP_ARMRU	P13029 escherichia
28	111.5	6.9	361	CCPR_ARMRU	P00431 saccharomyc
29	111.5	6.9	727	CATP_ARMRU	P17750 salmonella
30	96.5	6.0	382	PERM_ARMRU	P19136 phanerocae
31	94.5	5.9	1496	CPSM_ARMRU	P01233 rana catesb
32	93.5	5.8	1274	BXF_ARMRU	P30996 clostridium
33	92	5.7	463	FLGE_ARMRU	P056326 treponema p

34	92	5.7	1500	1	CPSM_HUMAN	P31327 homo sapien
35	90.5	5.6	244	1	FIMB_BOPE	P33409 bordetella
36	90	5.6	581	1	FIRB_ADE05	P11818 human adeno
37	90	5.6	1500	1	CPSM_RAT	P07756 rattus norv
38	89.5	5.6	468	1	ALSD_CANAL	P074657 candida alb
39	88	5.5	372	1	LIG6_PHACH	P50622 phanerocae
40	88	5.5	2009	1	SEC7_YEAST	P11075 saccharomyc
41	87.5	5.4	361	1	LIG_PHLRA	P20010 phlebia rad
42	86.5	5.4	763	1	ECHA_PIG	P029554 sus scrofa
43	86	5.3	647	1	COAT_ADYG	P24029 aleutian m1
44	85.5	5.3	275	1	YEIP_ECOLI	P33028 escherichia
45	85.5	5.3	682	1	HTFA_HUMAN	P09081 homo sapien

## ALIGNMENTS

RESULT	1	PERA_ARMRU	STANDARD;	PRT;	353 AA.
ID	PERA_ARMRU	P00433			
AC	21-JUL-1986 (Rel. 01, Created)				
DT	01-APR-1990 (Rel. 14, Last sequence update)				
DT	15-FEB-2000 (Rel. 39, Last annotation update)				
DE	PEROXIDASE C1A PRECURSOR (EC 1.11.1.7).				
GN	PRXC1A OR HPRC1.				
OS	Armoracia rusticana (Horseradish) (Armoracia lappatifolia).				
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;				
OC	Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;				
OC	Brassicaceae; Armoracia.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE: 88225087.				
RA	Fujiyama K., Takemura H., Shibayama S., Kobayashi K., Choi J.K.,				
RA	Shimizu A., Takano M., Yamada T., Okada H.;				
RT	"Structure of the horseradish peroxidase isozyme C genes.";				
RL	Eur. J. Biochem. 173:681-687(1988).				
RN	[2]				
RP	SEQUENCE OF 31-338.				
RX	MEDLINE: 77068850.				
RA	Wellinder K.G.;				
RT	"Covalent structure of the glycoprotein horseradish peroxidase (EC				
RL	1.11.1.7)."				
RN	FEBS Lett. 72:19-23(1976).				
RP	[3]				
RX	X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS).				
RA	Gajhede M., Schuller D.J., Henriksen A., Smith A.T., Poulos T.L.;				
RT	"Crystal structure of horseradish peroxidase C at 2.15-A resolution.";				
RL	Nat. Struct. Biol. 4:1032-1038(1997).				
RN	[4]				
RP	X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).				
RX	MEDLINE: 98069652.				
RA	Gajhede M., Schuller D.J., Henriksen A., Smith A.T., Poulos T.L.;				
RT	"Structural interactions between horseradish peroxidase C and the				
RL	substrate benzhydroxamic acid determined by x-ray crystallography.";				
RN	Biochemistry 37:8054-8060(1998).				
RP	[5]				
RX	X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).				
RA	Meno K., White C.G., Smith A.T., Gajhede M.;				
RT	Submitted (DEC-1998) to the PDB data bank.				
RL	"FUNCTION. REMOVAL OF H(2)O(2). OXIDATION OF TOXIC REDUCTANTS,				
CC	BIOSYNTHESIS AND DEGRADATION OF LIGNIN. DEFENSE RESPONSE TOWARD				
CC	WOUNDING AND METABOLISM OF AUXIN. THESE FUNCTIONS MIGHT BE				
CC	DEPENDENT ON EACH ISOZYME/ISOFORM IN EACH PLANT TISSUE.				
CC	-1- COFACTOR. BINDS HEME.				
CC	-1- SUBUNIT. MONOMER.				
CC	-1- SIMILARITY. HIGH, TO OTHER PLANT PEROXIDASES.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL, M37156; AAA33377.1; ALT\_SEQ.

DR PIR; A00502; OPRHC.

DR PIR; S00625; S00625.

DR PDB; 1ATJ; 04-FEB-98.

DR PDB; 2ATJ; 28-JAN-98.

DR PDB; 3ATJ; 23-DEC-98.

DR PFAM; PF00141; peroxidase; 1.

DR PRINTS; PR00458; PEROXIDASE.

DR PRINTS; PR00461; PLPEROXIDASE.

DR PROSITE; PS00435; PEROXIDASE\_1; 1.

DR PROSITE; PS00436; PEROXIDASE\_2; 1.

KM Oxidoreductase; Glycoprotein; Peroxidase; Heme; Multigene family;

KW Signal; 3D-structure.

FT SIGNAL 1 30

FT CHAIN 31 338 PEROXIDASE CIA.

FT PROPER 339 353 PYRROLIDONE CARBOXYLIC ACID.

FT MOD\_RES 31 31

FT ACT\_SITE 68 68

FT ACT\_SITE 72 72

FT ACT\_SITE 200 200

FT ACT\_SITE 41 121

FT DISULFID 74 79

FT DISULFID 127 331

FT DISULFID 207 239

FT CARBOHYD 43 43

FT CARBOHYD 87 87

FT CARBOHYD 188 188

FT CARBOHYD 216 216

FT CARBOHYD 228 228

FT CARBOHYD 244 244

FT CARBOHYD 285 285

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

FT CARBOHYD 298 298

AC P24101;

DT 01-MAR-1992 (Rel. 21, Created)

DT 01-MAR-1992 (Rel. 21, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE NEUTRAL PEROXIDASE C PRECURSOR (EC 1.11.1.7).

GN PRXCA.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;

OC Brassicaceae; Arabidopsis.

OC [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 91200671.

RA Intapur C., Higashimura N., Yamamoto K., Okada N., Shimmyo A.,

RA Takano M.;

RT "Nucleotide sequences of two genomic DNAs encoding peroxidase of

Arabidopsis thaliana."

RL Gene 98:237-241(1991).

CC -1- FUNCTION: REMOVAL OF H(2)O(2), OXIDATION OF TOXIC REDUCTANTS,

CC BIOSYNTHESIS AND DEGRADATION OF LIGNIN, DEFENSE RESPONSE TOWARD

CC WOUNDING AND METABOLISM OF AUXIN. THESE FUNCTIONS MIGHT BE

CC DEPENDENT ON EACH ISOZYME/ISOPORM IN EACH PLANT TISSUE.

CC -1- CATALYTIC ACTIVITY: DONOR + H(2)O(2) -> OXIDIZED DONOR + 2 H(2)O.

CC -1- COFACTOR: BINDS HEME.

CC -1- TISSUE SPECIFICITY: ROOTS.

CC -1- SIMILARITY: HIGH, TO OTHER PLANT PEROXIDASES.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL, M58380; AAA32849.1; -

DR PIR; J00457; J00457.

DR HSSP; P00433; 2ATJ.

DR PFAM; PF00141; peroxidase; 1.

DR PRINTS; PR00458; PEROXIDASE.

DR PRINTS; PR00461; PLPEROXIDASE.

DR PROSITE; PS00435; PEROXIDASE\_1; 1.

DR PROSITE; PS00436; PEROXIDASE\_2; 1.

KM Oxidoreductase; Glycoprotein; Peroxidase; Heme; Multigene family;

KW Signal.

FT SIGNAL 1 21

FT CHAIN 22 354

FT ACT\_SITE 69 69

FT ACT\_SITE 73 73

FT ACT\_SITE 201 201

FT DISULFID 42 122

FT DISULFID 75 80

FT DISULFID 128 332

FT DISULFID 208 240

FT DISULFID 288 332

FT CARBOHYD 88 88

FT CARBOHYD 217 217

FT CARBOHYD 229 229

FT CARBOHYD 245 245

FT CARBOHYD 286 286

FT CARBOHYD 286 286

FT CARBOHYD 286 286

FT CARBOHYD 286 286

FT CARBOHYD 286 286

FT CARBOHYD 286 286

FT CARBOHYD 286 286

FT CARBOHYD 286 286

FT CARBOHYD 286 286

FT CARBOHYD 286 286

FT CARBOHYD 286 286

FT CARBOHYD 286 286

FT CARBOHYD 286 286

FT CARBOHYD 286 286

FT CARBOHYD 286 286

FT CARBOHYD 286 286

FT CARBOHYD 286 286

Query Match

Best local Similarity 91.7%; Score 1475; DB 1; Length 354;

Matches 280; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

QY 2 QLTPTFYDSCPNVSNIVRDTIVNELRSDPRIASILRLHFHDCFVNGCDASILLDNNTS 61

DB 32 QLTPTFYDSCPTVNTIVRDTIVNELRSDPRIAGSILRLHFHDCFVNGCDASILLDNNTS 91

QY 62 FRTKEKAFGNANSARGFPVIDRKAVERACPRTVSCADMLTIAAQSVTLAGGSPWVRP 121

DB 92 FRTKEKALGNANSARGFPVIDRKAVERACPRTVSCADMLTIAAQSVTLAGGSPWVRP 151

FT	DISULFID	186	218	BY SIMILARITY.
FT	CARBOHYD	22	22	POTENTIAL.
FT	CARBOHYD	66	66	POTENTIAL.
FT	CARBOHYD	195	195	POTENTIAL.
FT	CARBOHYD	207	207	POTENTIAL.
FT	CARBOHYD	223	223	POTENTIAL.
FT	CARBOHYD	264	264	POTENTIAL.
SO	SEQUENCE	332 AA;	36548 MW;	1938A450D595DEBE CRC64;

  

Query Match	Best Local Similarity	91.6%;	Score 1473;	DB 1;	Length 332;
Matches 280;	Conservative 13;	Mismatches 15;	Indels 0;	Gaps 0;	

  

QY	2	QUTPEFDNCPNVSNIYRDTIVNELRSDPRIASILRLHFHDCFCVNGCDASILLDNMTS	61
Db	10	QUTPEFDNCPNVSNIYRDTIVNELRSDPRIASILRLHFHDCFCVNGCDASILLDNMTS	69
QY	62	FRTKDAAGNANSARGPVVDIRKAAVSAAPRVSADLTTLTAQOSVTLAAGPSMRVP	121
Db	70	FRTKDAAGNANSARGPVVDIRKAAVSAAPRVSADLTTLTAQOSVTLAAGPSMRVP	129
QY	122	LGRDLSLOAFDLDLANANLPAFPTLPOLKDSFRVNGVLRSSDIALVSGHFGKNCRFI	181
Db	130	LGRDLSLOAFDLDLANANLPAFPTLPOLKDSFRVNGVLRSSDIALVSGHFGKNCRFI	189
QY	182	MDRLYNESNTGLPDPPTLTLYTLQTLRGICPLNGNLSALVDFDLPTPTFDNRKYVNEEQ	241
Db	190	MDRLYNESNTGLPDPPTLTLYTLQTLRGICPLNGNLSALVDFDLPTPTFDNRKYVNEEQ	249
QY	242	KGLIOSDQELFSSPDATDTIPLVRSFANSSTOTFFNAFVPEANDRGNITPLTGQOIRLN	301
Db	250	KGLIOSDQELFSSPDATDTIPLVRSFANSSTOTFFNAFVPEANDRGNITPLTGQOIRLN	309
QY	302	CRVYNSNS	309
Db	310	CRVYNSNS	317

  

PERB_ARMRU	STANDARD;	PRT;	351 AA.
AC	P15232;		
DT	01-APR-1990 (Rel. 14, Created)		
DT	01-APR-1990 (Rel. 14, Last sequence update)		
DT	15-FEB-2000 (Rel. 39, Last annotation update)		
DE	PEROXIDASE C1B PRECURSOR (EC 1.11.1.7).		
CN	PRXC1B OR HPXC2.		
OS	Armoreria rusticana (Horse-radish) (Armoreria laphatfolia).		
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;		
OC	Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;		
OC	Brassicaceae; Armoreria.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	MEDLINE; 88225087.		
RA	Fujiyama K., Takemura H., Shibayama S., Kobayashi K., Choi J.K.,		
RA	Shiomi A., Takano M., Yamada Y., Okada H.;		
RT	"Structure of the horseradish peroxidase Isozyme C genes."		
RL	Eur. J. Biochem. 173:681-687(1988).		
CC	-1- FUNCTION: REMOVAL OF H(2)O(2), OXIDATION OF TOXIC REDUCTANTS,		
CC	BIOSYNTHESIS AND DEGRADATION OF LIGNIN, DEFENSE RESPONSE TOWARD		
CC	WOUNDING AND METABOLISM OF AUXIN. THESE FUNCTIONS MIGHT BE		
CC	DEPENDENT ON EACH ISOZYME/ISOFORM IN EACH PLANT TISSUE.		
CC	-1- CATALYTIC ACTIVITY: DONOR + H(2)O(2) -> OXIDIZED DONOR + 2 H(2)O.		
CC	-1- COFACTOR: BINDS HEME.		
CC	-1- SIMILARITY: HIGH, TO OTHER PLANT PEROXIDASES.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		

CC or send an email to license@lsb-sib.ch).  
-----  
CC EMBL; M37157; AAA33378.1; -.  
DR PIR; S00626; S00626.  
DR HSSP; P00433; 2ATU.  
DR PFAM; PF00141; peroxidase\_1.  
DR PRINTS; PR00458; peroxidase.  
DR PRINTS; PR00461; PLPEROXIDASE.  
DR PROSITE; PS00435; PEROXIDASE\_1; 1.  
DR PROSITE; PS00436; PEROXIDASE\_2; 1.  
KM Oxidoreductase; Glycoprotein; Peroxidase; Heme; Multigene family;  
KW Signal.  
FT SIGNAL 1 28  
FT CHAIN 29 351  
FT MOD\_RES 29 29  
FT ACT\_SITE 66 66 PEROXIDASE C1B.  
FT ACT\_SITE 70 70 PYRROLIDONE CARBOXYLIC ACID (BY  
FT ACT\_SITE 198 198 SIMILARITY).  
FT DISULFID 39 119 BY SIMILARITY.  
FT DISULFID 72 77 PROXIMAL HISTIDINE (HEME AXIAL LIGAND).  
FT DISULFID 125 329 BY SIMILARITY.  
FT DISULFID 205 237 BY SIMILARITY.  
FT CARBOHYD 41 41 POTENTIAL.  
FT CARBOHYD 85 85 POTENTIAL.  
FT CARBOHYD 214 214 POTENTIAL.  
FT CARBOHYD 226 226 POTENTIAL.  
FT CARBOHYD 242 242 POTENTIAL.  
FT CARBOHYD 283 283 POTENTIAL.  
SQ SEQUENCE 351 AA; 38645 MW; 7A8C606A3928950B CRC64;

Query Match 91.4%; Score 1469; DB 1; Length 351;  
Best local Similarity 90.6%; Pred. No. 2.5e-116;  
Matches 279; Conservative 12; Mismatches 17; Indels 0; Gaps 0;

QY 2 QLTPTFYDNCSPVNSIVRDTIVNELSDPRIASILRLHFHCFVNGCDASILLDMTTS 61  
DB 29 QLTPTFYDNCSPVNSIVRDTIVNELSDPRIASILRLHFHCFVNGCDASILLDMTTS 88  
QY 62 FTEKDAFGNANSARGFVIDRMKAVERACPTVSCADLLTTIAOOSVTLAGGSPMRVP 121  
DB 89 FTEKDAFGNANSARGFVIDRMKAVERACPTVSCADLLTTIAOOSVTLAGGSPMRVP 148  
QY 122 LGRDLSLOAFILDANANLPAFFFTLPOLKDSFRVNGLNSSDVLVALSGHTFGKNOCRFI 181  
DB 149 LGRDLSLOAFILDANANLPAFFFTLPOLKDAFAKVGGLDRSDVLVALSGHTFGKNOCRFI 208  
QY 182 MDRLYNSNGLPDPTLNTTYLQTLRGCLPLNGNLSALVDFDLRTPTIFDNKYYVNLREQ 241  
DB 209 MDRLYNSNGLPDPTLNTTYLQTLRGCLPLNGNLSALVDFDLRTPTIFDNKYYVNLREQ 268  
QY 242 KGLIOSDOELFSSPDATDTIPLVRSFANSQTGFENAFVEMDMRGNTITLTGTGGQIRLN 301  
DB 269 KGLIOSDOELFSSPDATDTIPLVRSFADGTQKFENAFVEMDMRGNTITLTGTGGQIRLN 328  
QY 302 CRYVNSNS 309  
DB 329 CRYVNSNS 336  
RESULT 5  
PER2\_ARMRU  
ID PER2\_ARMRU STANDARD: PRT: 347 AA.  
AC P17179;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 15-FEB-2000 (Rel. 39, Last annotation update)  
DE PEROXIDASE C2 PRECURSOR (EC 1.11.1.7).  
GN PRXC2.  
OS Armoracia rusticana (Horse radish) (Armoracia laphactifolia).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;

CC Brassicaceae; Armoracia.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 90323613.  
RA Fujiyama K., Takemura H., Shimmyo A., Okada H., Takano M.;  
RT "Genomic DNA structure of two new horseradish-peroxidase-encoding  
genes";  
RL Gene 89:163-169(1990).  
CC -1- FUNCTION: REMOVAL OF H(2)O(2), OXIDATION OF TOXIC REDUCTANTS,  
CC BIOSYNTHESIS AND DEGRADATION OF LIGNIN, DEFENSE RESPONSE TOWARD  
CC WOUNDING AND METABOLISM OF AUXIN. THESE FUNCTIONS MIGHT BE  
CC DEPENDENT ON EACH ISOZYME/ISOFORM IN EACH PLANT TISSUE.  
CC -1- CATALYTIC ACTIVITY: DONOR + H(2)O(2) = OXIDIZED DONOR + 2 H(2)O.  
CC -1- COFACTOR: BINDS HEME.  
CC -1- SIMILARITY: HIGH, TO OTHER PLANT PEROXIDASES.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.lsb-sib.ch/announce/>  
CC or send an email to license@lsb-sib.ch).

CC EMBL; D90115; BAA14143.1; -.  
DR PIR; JH0149; JH0149.  
DR HSSP; P00433; 2ATU.  
DR PFAM; PF00141; peroxidase\_1.  
DR PRINTS; PR00458; peroxidase.  
DR PRINTS; PR00461; PLPEROXIDASE.  
DR PROSITE; PS00435; PEROXIDASE\_1; 1.  
DR PROSITE; PS00436; PEROXIDASE\_2; 1.  
KM Oxidoreductase; Glycoprotein; Peroxidase; Heme; Multigene family;  
KW Signal.  
FT SIGNAL 1 24  
FT CHAIN 25 347  
FT MOD\_RES 25 25  
FT ACT\_SITE 62 62 PEROXIDASE C2.  
FT ACT\_SITE 66 66 PYRROLIDONE CARBOXYLIC ACID (BY  
FT ACT\_SITE 194 194 SIMILARITY).  
FT DISULFID 35 115 DISTAL HISTIDINE (BY SIMILARITY).  
FT DISULFID 68 73 PROXIMAL HISTIDINE (HEME AXIAL LIGAND).  
FT DISULFID 121 325 BY SIMILARITY.  
FT DISULFID 201 233 BY SIMILARITY.  
FT CARBOHYD 81 81 POTENTIAL.  
FT CARBOHYD 210 210 POTENTIAL.  
FT CARBOHYD 238 238 POTENTIAL.  
SQ SEQUENCE 347 AA; 38035 MW; 3EE9A2CFDECA49A CRC64;

Query Match 79.0%; Score 1271; DB 1; Length 347;  
Best local Similarity 79.4%; Pred. No. 1.1e-99;  
Matches 243; Conservative 24; Mismatches 39; Indels 0; Gaps 0;

QY 2 QLTPTFYDNCSPVNSIVRDTIVNELSDPRIASILRLHFHCFVNGCDASILLDMTTS 61  
DB 25 QLTPTFYDNCSPVNSIVRDTIVNELSDPRIASILRLHFHCFVNGCDASILLDMTTS 84  
QY 62 FTEKDAFGNANSARGFVIDRMKAVERACPTVSCADLLTTIAOOSVTLAGGSPMRVP 121  
DB 85 FTEKDAFGNANSARGFVIDRMKAVERACPTVSCADLLTTIAOOSVTLAGGSPMRVP 144  
QY 122 LGRDLSLOAFILDANANLPAFFFTLPOLKDSFRVNGLNSSDVLVALSGHTFGKNOCRFI 181  
DB 145 LGRDLSLOAFILDANANLPAFFFTLPOLKDSFRVNGLNSSDVLVALSGHTFGKNOCRFI 204  
QY 182 MDRLYNSNGLPDPTLNTTYLQTLRGCLPLNGNLSALVDFDLRTPTIFDNKYYVNLREQ 241  
DB 205 MDRLYNSNGLPDPTLNTTYLQTLRGCLPLNGNLSALVDFDLRTPTIFDNKYYVNLREQ 264  
QY 242 KGLIOSDOELFSSPDATDTIPLVRSFANSQTGFENAFVEMDMRGNTITLTGTGGQIRLN 301  
DB 269 KGLIOSDOELFSSPDATDTIPLVRSFADGTQKFENAFVEMDMRGNTITLTGTGGQIRLN 328

Qy	2	QUTPTFDNSCPNPNSTIVROTIYVELSDSPRIASILRLRHDOCFVNCGDASILDNSTS	61
Dd	30	QLRDFEFTTCTPSVFNNIIGDIIVDELRTDPRIIASLMLHFHDFVRCCDASILDNSTS	89
Qy	62	FRTKEDAFGNANSARGPEVIDIRMKAAVASACPRTVSCADLLITIAAQSQSVTLAAGSMPWVP	121
Dd	90	FRTKEDAPRNANSARGVEVIDRMKTSLERACPRVVSADVLTITASISVLSGGPMWFVP	149
Qy	122	LGRDSIOALFLDLANANLPAPFPLLPOLKDSFRVNGLRSSDLVALSGCHTFGKNOCRFI	181
Dd	150	LGRDSVEAFEDLDNTALPSEFTLIAQKKAFADYGLNRPSDVLALSCHTFEGRAQCFFV	209
Qy	182	MDRLYNSNTGLPPPTLNTTYTLQTLRGCLPLANGLSALVDVDTLPTTFDKNKYYNLEEQ	241
Dd	210	TPLRYNFNGTRRPPTIDPYTLVOLRALCPCPNGMGTVLVNDVYTPNPFDEQOYTNLNKG	269
Qy	242	KGLIOSDOELSSPDATPTIFLVRSFANSSTOTFFNAFEANDRMGINIPLGTOGOURLN	301
Dd	270	KGLIOSDOELSTFGCA-DTIFLVNLYSNSTTAFFGAFAVDAMIRKMNLRLPGTGOELNON	328
Qy	302	CRVYNS 307	
Dd	329	CRVYNS 334	
RESULT	7		
PERE_ARATH	ID	STANDARD;	PRT; 349 AA.
AC	P24102:		
DT	01-MAR-1992 (Rel. 21, Created)		
DT	01-MAR-1992 (Rel. 21, Last sequence update)		
DT	01-OCT-1996 (Rel. 34, Last annotation update)		
DE	BASIC PEROXIDASE E PRECURSOR (EC 1.11.1.7).		
GN	PREVA.		
OC	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;		
OC	Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;		
OC	Brassicaceae; Arabidopsi-		
RN	[1]		
RA	SEQUENCE FROM N.A.		
RX	MEDLINE; G1200671.		
RA	Inapuk C., Higashimura N., Yamamoto K., Okada N., Shimmyo A.,		
RT	Takano M.;		
RT	"Nucleotide sequences of two genomic DNAs encoding peroxidase of		
RL	Arabidopsis thaliana.";		
CC	Gene 98:237-241(1991).		
CC	-1- FUNCTION: REMOVAL OF H(2)O(2), OXIDATION OF TOXIC REDUCTANTS,		
CC	BIOSYNTHESIS AND DEGRADATION OF LIGNIN, DEFENSE RESPONSE TOWARD		
CC	WOUNDING AND METABOLISM OF AUXIN. THESE FUNCTIONS MIGHT BE		
CC	DEPENDENT ON EACH ISOZYME/ISOFORM IN EACH PLANT TISSUE.		
CC	-1- CATALYTIC ACTIVITY: DONOR + H(2)O(2) -> OXIDIZED DONOR + 2 H(2)O.		
CC	-1- COFACTOR: BINDS HEME.		
CC	-1- TISSUE SPECIFICITY: ROOTS.		
CC	-1- SIMILARITY: HIGH, TO OTHER PLANT PEROXIDASES.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
DR	EMBL; MS8381; AAA32842.1; .		
DR	PIR; J00458; J00458.		
DR	HSSP; P00433; IATU.		
DR	PFAM; PF00141; peroxidase_1.		
DR	PRINTS; PR00458; PEROXIDASE.		
DR	PRINTS; PR00461; PUEROXIDASE.		
DR	PROSITE; PS00435; PEROXIDASE_1; 1.		
DR	PROSITE; PS00436; PEROXIDASE_2; 1.		
DR	Oxidoreductase; Glycoprotein; Peroxidase; Heme; Multigene family;		
FW			





CC -1- SIMILARITY: HIGH, TO OTHER PLANT PEROXIDASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: J02979; AAA34108.1; -  
DR PIR: A39889; A39889.  
DR HSSP: P00433; 1ATU.  
DR PFAM: PF00141; peroxidase; 1.  
DR PRINTS: PR00458; PEROXIDASE.  
DR PRINTS: PR00461; PLPEROXIDASE.  
DR PROSITE: PS00435; PEROXIDASE\_1; 1.  
DR PROSITE: PS00436; PEROXIDASE\_2; 1.  
KW Oxidoreductase; Peroxidase; Heme; Glycoprotein; Signal.  
FT SIGNAL 1 22  
FT CHAIN 23 324 LIGNIN FORMING ANIONIC PEROXIDASE.  
FT MOD\_RES 23 23 PYRROLIDONE CARBOXYLIC ACID.  
FT ACT\_SITE 60 60 BY SIMILARITY.  
FT ACT\_SITE 64 64 DISTAL HISTIDINE (BY SIMILARITY).  
FT ACT\_SITE 189 189 PROXIMAL HISTIDINE (HEME AXIAL LIGAND)  
FT (BY SIMILARITY).  
FT DISULFID 33 111 BY SIMILARITY.  
FT DISULFID 66 71 BY SIMILARITY.  
FT DISULFID 117 320 BY SIMILARITY.  
FT DISULFID 196 228 BY SIMILARITY.  
FT CARBOHYD 35 35 POTENTIAL.  
FT CARBOHYD 150 150 POTENTIAL.  
FT CARBOHYD 207 207 POTENTIAL.  
SQ SEQUENCE 324 AA; 34674 MW; 0F1F03927F47A669 CRC64;

Query Match 48.9%; Score 786.5; DB 1; Length 324;  
Best Local Similarity 51.0%; Pred. No. 5.1e-59;  
Matches 156; Conservative 56; Mismatches 89; Indels 5; Gaps 5;  
OY 2 QLTPTFYDNCSPNVSNIYRDTIVNELSDPRIASILRLHDFCVNGCDASILLDTTS 61  
DB 23 QLSATPFYDTTCNPVTSIVRGVMDORORDARAGAKIIRLHPHDFVNGCDSILLD-TDG 81  
OY 62 FRTKEDAFGNANSARGFVIDIRMAKAVESACPTVSCADLLTTIAAGSVTLAGGSPWRP 121  
DB 82 TOTKEDAPANV-GAGGFDIVDITALENCPGVASCADILALSELGVVLAKGSPWVL 140  
OY 122 LGRDLSLOAFILANANLPAPFTLPLKDSFRVNGLRSSDVALSGGHTFGKNQCFI 181  
DB 141 EGRDLSLANSNGANSNDIPSPFETLAVIPIQFTKGM-DLTDVALSGAHFGARCTGF 199  
OY 182 MDRLYNSNGLPPTLTNTTYLQTLKGLCPNLGNL-SALVDFDLRTPIIFNKYYVLEE 240  
DB 200 EQRLENFNGSCNPULTYDAFELQTLGICPOGGNNGNFTMLDISTPNDYFTNLQS 259  
OY 241 OKGLISDOELFSSPDATDRTPLVRSFANSQTGFENAFVEAMGNITPLTGTGOGRL 300  
DB 260 NQGLQTDDELFT-SSGATIAVNRVAGSOTOFFDFVSMKGNISPLGTNGQIRT 318  
OY 301 NCRVYN 306  
DB 319 DCKRVN 324

RESULT 10  
PERL\_ARAHY STANDARD; PRT; 316 AA.  
AC P2195;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE CATIONIC PEROXIDASE 1 PRECURSOR (EC 1.11.1.7).

GN PNC1.  
OS *Arachis hypogaea* (Peanut).  
OC Eukaryota: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta: eudicotyledons; Rosidae; eurosids I; Fabales; Fabaceae;  
OC Papilionoideae; Arachis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 91062381.  
RA Buffard D., Breda C., van Huystee R.B., Asemota O., Pierre M.,  
RA Dang Ha D.B., Esnault R.;  
RT "Molecular cloning of complementary DNAs encoding two cationic  
RT peroxidases from cultivated peanut cells.";  
RL Proc. Natl. Acad. Sci. U.S.A. 87:8874-8878(1990).  
RN [2]  
RP REVISION TO 47.  
RA Esnault R.;  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).  
RX MEDLINE; 96398617.  
RA Schuller D.J., Ban N., van Huystee R.B., McPherson A., Poulos T.L.;  
RT "The crystal structure of peanut peroxidase.";  
RL Structure 4:311-321(1996).  
CC -1- FUNCTION: REMOVAL OF H(2)O(2), OXIDATION OF TOXIC REDUCTANTS,  
CC BIOSYNTHESIS AND DEGRADATION OF LIGNIN, DEFENSE RESPONSE TOWARD  
CC WOUNDING AND METABOLISM OF AUXIN. THESE FUNCTIONS MIGHT BE  
CC DEPENDENT ON EACH ISOZYME/ISOFORM IN EACH PLANT TISSUE.  
CC -1- CATALYTIC ACTIVITY: DONOR + H(2)O(2) = OXIDIZED DONOR + 2 H(2)O.  
CC -1- COFACTOR: BINDS HEME.  
CC -1- SIMILARITY: HIGH, TO OTHER PLANT PEROXIDASES (CLASS III).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration -  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M37636; AAB06183.1; -  
DR PIR: A38265; A38265.  
DR PDB: 1SCH; 1I-JUL-96.  
DR PFAM: PF00141; peroxidase; 1.  
DR PRINTS: PR00458; PEROXIDASE.  
DR PRINTS: PR00461; PLPEROXIDASE.  
DR PROSITE: PS00435; PEROXIDASE\_1; 1.  
DR PROSITE: PS00436; PEROXIDASE\_2; 1.  
KW Oxidoreductase; Glycoprotein; Peroxidase; Heme; Multigene family;  
KW Signal; 3D-structure.  
FT SIGNAL 1 22 POTENTIAL.  
FT CHAIN 23 316 CATIONIC PEROXIDASE 1.  
FT MOD\_RES 23 23 PYRROLIDONE CARBOXYLIC ACID.  
FT ACT\_SITE 60 60 BY SIMILARITY.  
FT ACT\_SITE 64 64 DISTAL HISTIDINE (BY SIMILARITY).  
FT ACT\_SITE 191 191 PROXIMAL HISTIDINE (HEME AXIAL LIGAND).  
FT CARBOHYD 82 82 POTENTIAL.  
FT CARBOHYD 166 166 POTENTIAL.  
FT CARBOHYD 207 207 POTENTIAL.  
FT DISULFID 33 113 POTENTIAL.  
FT DISULFID 66 71  
FT DISULFID 119 312  
FT DISULFID 198 223  
SQ SEQUENCE 316 AA; 33517 MW; 2CC271F8B8B8C9F0 CRC64;

Query Match 46.4%; Score 745.5; DB 1; Length 316;  
Best Local Similarity 49.2%; Pred. No. 1.4e-55;  
Matches 150; Conservative 45; Mismatches 99; Indels 11; Gaps 3;  
OY 2 QLTPTFYDNCSPNVSNIYRDTIVNELSDPRIASILRLHDFCVNGCDASILLDTTS 61  
DB 23 QLSATPFYDTTCNPVTSIVRGVMDORORDARAGAKIIRLHPHDFVNGCDSILLDTTSN 82

QY 62 FTEKDAFGNANSARGPEVIDRMKAIVESACPTVSCADLLTTAAGOSVTLAGPSMRVP 121  
 DB 83 FTEGEKTGPNVANSIRGEVITIKSQVESICPGVYSCADILAAVARSVALGASMMVL 142  
 QY 122 LGRDSTLQATLDLANANLPAPEFTLPOLKDSFRVNGLNRRSSDLVALSGHTEGNCQRFI 181  
 DB 143 LGRDSTTALSSANSNDLPAPFENLGLISAFSNKGF-TKELVATLSAHTIGAOCTAP 201  
 QY 182 MDRLYFNSNTGLPDPPTLNTTYLTQTLRGICPLNGNLSALVDFDLRTPTIFDNKYYVNLREQ 241  
 DB 202 RTIRYNSN-----IDPTTAKSLQANCPSVGGDTLNSPFDVTPPKFDMAYINLRNK 254  
 QY 242 KGLIQSDQELFSSPDADTDTPLVRSFANSTQTFENAFVEMDRNGNTPLTGTGQIRLN 301  
 DB 255 KGLHSDQQLF-----NGVSTDSQVLTAYSNMATFNTDFGNAMIKMGNLSPLTGTSGQIRTN 311  
 QY 302 CRVYN 306  
 DB 312 CRKTN 316

RESULT 11

PER2\_CUCSA STANDARD; PRT; 292 AA.  
 ID PER2\_CUCSA

AC P19135;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 01-MAR-1992 (Rel. 21, Last annotation update)  
 DE PEROXIDASE 2 (EC 1.11.1.7) (FRAGMENT).  
 OS Cucumis sativus (Cucumber).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Cucurbitales;  
 OC Cucurbitaceae; Cucumis.  
 RN [1]

SEQUENCE FROM N.A.

RX MEDLINE: 91346662.

RA Morgens P.H., Callahan A.M., Dunn L.J., Abeles F.B.;

RT Isolation and sequencing of cDNA clones encoding ethylene-induced

RL Plant Mol. Biol. 14:715-725(1990).

CC -1- CATALYTIC ACTIVITY: DONOR + H(2)O(2) = OXIDIZED DONOR + 2 H(2)O.

CC -1- COFACTOR: BINDS HEME.

CC -1- SIMILARITY: HIGH, TO OTHER PLANT PEROXIDASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL: M32742; AAA33121.1; -.

DR PIR: S11870; S11870.

DR HSSP: P00433; IATU.

DR PFAM: PF00141; peroxidase: 1.

DR PROSITE: PS00435; PEROXIDASE\_1: 1.

DR PROSITE: PS00436; PEROXIDASE\_2: 1.

KW Oxidoreductase; Peroxidase; Heme; Glycoprotein.

FT NON\_TER 1 1

FT ACT\_SITE 36 36 BY SIMILARITY.

FT ACT\_SITE 38 38 DISTAL HISTIDINE (BY SIMILARITY).

FT ACT\_SITE 164 164 PROXIMAL HISTIDINE (HEME AXIAL LIGAND).

FT CARBOHYD 68 68 POTENTIAL.

FT CARBOHYD 139 139 POTENTIAL.

FT CARBOHYD 179 179 POTENTIAL.

SQ SEQUENCE 292 AA; 31856 MW; 98FD89EF01CE6888 CRC64;

Query Match 45.9%; Score 737.5; DR 1; Length 292;  
 Best Local Similarity 52.8%; Pred. No. 5,8e-55;  
 Matches 159; Conservative 34; Mismatches 99; Indels 9; Gaps 6;

QY 6 TFDNSCPVNSIVRTIYNELRSDPRIAASILRLHDFCVNGCDAIILDNNTSFRTE 65  
 DB 1 TFDVESCDDVNSIRVRVQOALVSDERAGARLRLRHFHDFCVNGCDAIILDNNTSFRTE 60  
 QY 66 KDAFGNANSARGPEVIDRMKAIVESACPTVSCADLLTTAAGOSVTLAGPSMRVPYIGRR 125  
 DB 61 LAACGNAN-ITGFEIVANNIKAAVEKACPGVYSCADILAIASVSGVNLGAPCWCVOIGRR 119  
 QY 126 DSIQAFDLNANLPAPEFTLPOLKDSFRVNGLNRRSSDLVALSGHTEGNCQRFIDRL 185  
 DB 120 DSRANLQAGAIIDGLPSFEVNTOLKRFDRYDD-DTDLVALSGAHTFGKSCQF-FDRR 177  
 QY 186 YFNSNTGLPDPPTLNTTYLTQTLRGICPLNGNLSALVDFDLRTPTIFDNKYYVNLREQGLI 245  
 DB 178 LANSN--PDSTLNPRYAQQLRQAC--SSGRDFEVNLDPTTPPKFDMAYINLRNK 232  
 QY 246 QSDQELFSSPDADTDTPLVRSFANSTQTFENAFVEMDRNGNTPLTGTGQIRLCRV 305  
 DB 233 TSDQVHSTP-GEDTVKIVNLFASQNOFFESFGOSIMNGNIPILOGNGEIRSNCRRL 291  
 QY 306 N 306  
 DB 292 N 292

RESULT 12

PERL\_WHEAT STANDARD; PRT; 312 AA.  
 ID PERL\_WHEAT

AC Q05855;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 01-FEB-1995 (Rel. 31, Last annotation update)

DE PEROXIDASE PRECURSOR (EC 1.11.1.7).

OS Triticum aestivum (wheat).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; Liliopsida; Poales; Poaceae; Triticum.

RN [1]

RX SEQUENCE FROM N.A.

RC STRAIN-CV. CHEYENNE;

RX MEDLINE: 91363838.

RA Hertig C., Rehmann G., Bull J., Mauch F., Dudler R.;

RT "Sequence and tissue-specific expression of a putative peroxidase

RL Plant Mol. Biol. 16:171-174(1991).

CC -1- FUNCTION: INVOLVED IN DEFENSE RESPONSE TO POWDERY MELDREW FUNGUS.

CC -1- CATALYTIC ACTIVITY: DONOR + H(2)O(2) = OXIDIZED DONOR + 2 H(2)O.

CC -1- COFACTOR: BINDS HEME.

CC -1- SUBCELLULAR LOCATION: SECRETED (PROBABLE).

CC -1- TISSUE SPECIFICITY: ROOT.

CC -1- SIMILARITY: HIGH, TO OTHER PLANT PEROXIDASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL: X53675; CAA37713.1; -.

DR PIR: S13325; S13325.

DR HSSP: P22195; ISCH.

DR PFAM: PF00141; peroxidase: 1.

DR PRINTS: PR00458; PEROXIDASE.

DR PRINTS: PR00461; PEROXIDASE.

DR PROSITE: PS00435; PEROXIDASE\_1: 1.

DR PROSITE: PS00436; PEROXIDASE\_2: 1.

KW Oxidoreductase; Peroxidase; Heme; Glycoprotein; Multigene family;

FT SIGNAL 1 23

FT CHAIN 24 312 PEROXIDASE.

FT ACT\_SITE 61 61 BY SIMILARITY.

FT ACT\_SITE 65 65 DISTAL HISTIDINE (BY SIMILARITY).



CC solanaceae; Nicotiana.  
CC [1]  
CC SEQUENCE FROM N.A.  
CC STRAIN=CV. XANTHI NC;  
CC MEDLINE: 93041285.  
CC Criqui M.-C., Plesse B., Durr A., Marbach J., Parmentier Y.,  
CC Jamet E., Fleck J.;  
CC "characterization of genes expressed in mesophyll protoplasts of  
CC Nicotiana glauca before the re-initiation of the DNA  
CC replicational activity.";  
CC Mech. Dev. 38:121-132(1992).  
CC -1- FUNCTION: THIS ENZYME PLAYS AN INTEGRAL ROLE IN SECONDARY CELL  
CC WALL BIOSYNTHESIS BY THE POLYMERIZATION OF CINNAMYL ALCOHOLS,  
CC INTO LIGNIN AND BY FORMING RIGID CROSS-LINKS BETWEEN CELLULOSE,  
CC PECTIN, HYDROXY-PROLINE-RICH GLYCOPROTEINS, AND LIGNIN.  
CC -1- CATALYTIC ACTIVITY: DONOR + H(2O)(2) = OXIDIZED DONOR + 2 H(2O).  
CC -1- COFACTOR: BINDS HEME.  
CC -1- TISSUE SPECIFICITY: MESOPHYLL PROTOPLASTS AND TO A MUCH LESSER  
CC EXTENT, ROOTS AND GERMINATING SEEDS.  
CC -1- DEVELOPMENTAL STAGE: BEFORE RE-INITIATION OF THE DNA REPLICATIONAL  
CC ACTIVITY.  
CC -1- SIMILARITY: HIGH, TO OTHER PLANT PEROXIDASES.  
CC  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC

Query Match	Best Local Similarity	Matches 149; Conservative	43.7%;	Score 702;	DB 1;	Length 322;
Q1PPTFYDNCSPVNSIVRDTIVNELRSDPRIASILRLHFHDCFVNGCDASILLDDMTTS 61	48.9%;	42;	Pred. No. 6.3e-52;	Mismatches 104;	Indels 10;	Gaps 5;
Q1SSTFTDNTCPNALNTRISVKAISERRMAASLIRLHFHDCFVNGCDASILLDDTPS 87						
62 FRTKDAFGNANSARGPVIDIRMKAAVESACPTVSACADLLTTIAAQSVTLAGSPSMRVP 121						
88 IESKRTALPLMGARSARGGIIEDAKREVEKICPGVSCADILITVAARASAAVGPSPMTVK 147						
122 LGRDLSIQAFLDLANANLPAFPFTLLPOLKDSFRVNGLNRSDDVALSGGHTFGNOCRFI 181						
148 IGRDRSTASKTLAETLDPGPFDPDLNLTISFASKGS-TRDVALTSGAHTIGQAQCFLE 206						
182 MDRLYNNSNIGLFDPTLNTTYLQTLRGLCPNLGMSLAVDFDALTPTPIITFNKRYVNEEQ 241						

Db	207	RDRY--SN---GTDIDAGFASTRRROCPQEGENGNIAPLDLVTNPQFDNNYFKNLJOK	260
QY	242	KGLIQSDQELFSSPDATDTIPIVRSFANSTOTFENAFVEMDRMGNTTPLTGTGQIRLN	301
Db	261	KGLIQSDQVLENG-GSTDNI--VSEISNSARAFSSDFAAMIKMGDISPLSGONGIIRKV	317
QY	302	CRVYN	306
Db	318	CGSYN	322

Search completed: October 4, 2000, 13:04:57  
Job time: 1688 sec

**THIS PAGE BLANK (USPTO)**



```
Db 31 QLRPTFYDNCSPVNTIVRTIYNELRSDPRIASILRLHFHDCFVNGCDASILLDMTTS 90
QY 62 FRTKDAFGNANSARGPVIDIRKMAVESACPRTVSCADLLTTIAAOSVTLAGPSMRVP 121
    |||||
Db 91 FRTKDAFGNANSARGPVIDIRKMAVESACPRTVSCADLLTTIAAOSVTLAGPSMRVP 150
QY 122 LGRDLSQAFLDANANLPAFFFTLPOLKDSFRVNGINRSSDLVALSGHTEGKNCQCFI 181
    |||||
Db 151 LGRDLSQAFLDANANLPAFFFTLPOLKDSFRVNGINRSSDLVALSGHTEGKNCQCFI 210
QY 182 MDRLYNFSNTGLPDPPTLNTTYLQTLRGCLPLNGNLSALVDFDLRTPTIPDNKYYVNLREQ 241
    |||||
Db 211 LDRFYNSNTGLPDPPTLNTTYLQTLRGCLPLNGNRSALVDFDLRTPTIPDNKYYVNLKER 270
QY 242 KGLIQSDQELFSSPDATDTIPLVRSFANSQTQTFENAFVEMADMGNITPLTGTGQIRLN 301
    |||||
Db 271 KGLIQSDQELFSSPDATDTIPLVRSFANSQTQTFENAFVEMADMGNITPLTGTGQIRLN 330
QY 302 CRVYNSNS 309
    |||||
Db 331 CRVYNSNS 338

RESULT 2
Q43732 PRELIMINARY; PRT; 352 AA.
AC Q43732;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, last annotation update)
DE PEROXIDASE PRECURSOR (EC 1.11.1.7) (MELOPEROXIDASE).
GN YPR9 OR PRX3 OR PRX3.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VARIOUS STRAINS;
RA CAPELLI N., TOGNOLLI M., FLACH J., OVERNEY S., PENEL C., GREPPIN H.,
RA SIMON P.;
RN Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-VARIOUS STRAINS;
RA WESTERGAARD L., ABELSKOV A.K., JENSEN R.B., HANSEN L.N.,
RA RASMUSSEN S.K.;
RN Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TOGNOLLI M., GREPPIN H., SIMON P.;
RT "Structure of the gene encoding Arabidopsis thaliana PRX3
RT peroxidase."
RT Submitted (FEB-1999) to EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: DONOR + H(2)O(2) = OXIDIZED DONOR + 2 H(2)O.
CC -1- COFACTOR: HEME.
DR EMBL: X98315; CA66959.1; -
DR EMBL: X98777; CA67313.1; -
DR EMBL: A1133036; CAB37193.1; -
DR HSSP: P00433; ZATU.
DR MENDEL: 6566; Arabid. Ypr9.6566.
DR PROSITE: PS00435; PEROXIDASE_1; 1.
DR PROSITE: PS00436; PEROXIDASE_2; 1.
DR PFAM: PF00141; peroxidase: 1.
DR PRINTS: PR00458; PEROXIDASE.
DR PRINTS: PR00461; PLPEROXIDASE.
DR SIGNAL: 1 29 POTENTIAL.
FT CHAIN 30 337 POTENTIAL.
SQ SEQUENCE 352 AA; 38873 MW; 19DBBFDI CRC32;
```

```
Query Match 89.2%; Score 1434; DB 10; Length 352;
Best Local Similarity 88.6%; Pred. No. 2.5e-121;
Matches 273; Conservative 16; Mismatches 19; Indels 0; Gaps 0;

QY 2 QLRPTFYDNCSPVNSIVRDTIYNELRSDPRIASILRLHFHDCFVNGCDASILLDMTTS 61
    |||||
Db 30 QLRPTFYDNCSPVNTIVRTIYNELRSDPRIASILRLHFHDCFVNGCDASILLDMTTS 89
QY 62 FRTKDAFGNANSARGPVIDIRKMAVESACPRTVSCADLLTTIAAOSVTLAGPSMRVP 121
    |||||
Db 90 FRTKDAFGNANSARGPVIDIRKMAVESACPRTVSCADLLTTIAAOSVTLAGPSMRVP 149
QY 122 LGRDLSQAFLDANANLPAFFFTLPOLKDSFRVNGINRSSDLVALSGHTEGKNCQCFI 181
    |||||
Db 150 LGRDLSQAFLDANANLPAFFFTLPOLKDSFRVNGINRSSDLVALSGHTEGKNCQCFI 209
QY 182 MDRLYNFSNTGLPDPPTLNTTYLQTLRGCLPLNGNLSALVDFDLRTPTIPDNKYYVNLREQ 241
    |||||
Db 210 MDRLYNFSNTGLPDPPTLNTTYLQTLRGCLPLNGNQVNLVDFDLRTPTIPDNKYYVNLKEL 269
QY 242 KGLIQSDQELFSSPDATDTIPLVRSFANSQTQTFENAFVEMADMGNITPLTGTGQIRLN 301
    |||||
Db 270 KGLIQSDQELFSSPDATDTIPLVRSFANSQTQTFENAFVEMADMGNITPLTGTGQIRLN 329
QY 302 CRVYNSNS 309
    |||||
Db 330 CRVYNSNS 337

RESULT 3
O80912 PRELIMINARY; PRT; 349 AA.
ID O80912;
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, last annotation update)
DE PEROXIDASE.
GN YPR9 OR T19C21.12.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV, COLUMBIA;
RA ROUNSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,
RA SYKES S.M., KAUL S., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,
RA SOMERVILLE C.R., VENTER J.C.;
RN "Arabidopsis thaliana chromosome II BAC T19C21 genomic sequence.";
RN Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC004683; AAC28765.1; -
DR HSSP: P00433; IATU.
DR MENDEL: 31568; Arabid. Ypr9.31568.
DR PROSITE: PS00435; PEROXIDASE_1; 1.
DR PROSITE: PS00436; PEROXIDASE_2; 1.
DR PFAM: PF00141; peroxidase: 1.
DR PRINTS: PR00458; PEROXIDASE.
DR PRINTS: PR00461; PLPEROXIDASE.
KW Peroxidase.
SQ SEQUENCE 349 AA; 38100 MW; 8103AA13 CRC32;

Query Match 68.9%; Score 1108.5; DB 10; Length 349;
Best Local Similarity 70.5%; Pred. No. 4.9e-92;
Matches 215; Conservative 31; Mismatches 58; Indels 1; Gaps 1;

QY 2 QLRPTFYDNCSPVNSIVRDTIYNELRSDPRIASILRLHFHDCFVNGCDASILLDMTTS 61
    |||||
Db 30 QLRPDTFYRCPPIFNIGDTIYNELRSDPRIASILRLHFHDCFVNGCDASILLDMTTS 89
QY 62 FRTKDAFGNANSARGPVIDIRKMAVESACPRTVSCADLLTTIAAOSVTLAGPSMRVP 121
```



Db		90	FETEDDAPPNKNSVAGFVIDYDKMAIERACRYTSCADITIIAISOISLSSGHTFGMWVPV	149
Oy		122	LGRDSLOAFDLNANANI.PAPEFTLLPOLKDSFRNYGLNRSSDLVALSGGHTFGKNOCRTI	181
Db		150	LGRDSVEAFALANTWALPSPSTLTQKTARFADVGILNNPSDLVALSGGHTFGKNOCQCV	209
Oy		162	MRLTYNFSTGTLPDTLTNTTYIQTIRGLCPLNGNISALVDPLRPTTIIDNKYYVNLREQ	241
Db		210	TPLTNFMGNTPDPSLNPTYLVLELRRLCPONGNGTIVLVNFDSVPTPEDROQYTNLLNG	269
Oy		242	KGLISDOELFSSPATPTPIVRBSFANSSTOFENFAPEAMORMNITPLOTGOIRIN	301
Db		270	KGLISDOVLSTPCA-DTIPLVNOYSSTMTVFCAFDAMIRMKNLAKPLTGTOGEIRON	328
Oy		302	CRVVN	306
Db		329	CRVVN	333
RESULT		4		
O80913				
ID	PRELIMINARY:	PRT:	349 AA.	
AC	O80913:			
DT	01-NOV-1998 (TREMBLrel. 08, Created)			
DT	01-NOV-1998 (TREMBLrel. 08, Last sequence update)			
DT	01-NOV-1999 (TREMBLrel. 12, Last annotation update)			
DE	PEROXIDASE.			
GN	YPR9 OR T19C21.13.			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota, Viridiplantae, Streptophyta; Embryophyta; Tracheophyta;			
OC	euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;			
OC	core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;			
OC	Arabiadoptsis.			
[1]				
RP	SEQUENCE FROM N.A.			
RC	STRAIN-CV. COLUMBIA:			
RA	ROUNDEST S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,			
RA	STYES S.M., KAUL S., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,			
RA	SOMERVILLE C.R., VENTER J.C.;			
RT	"Arabidopsis thaliana chromosome II BAC T19C21 genomic sequence.";			
RL	Submitted (AUG-1998) to the EMBL/Genbank/DBJ databases.			
DR	HSSB; AC004683; AAC28766.1; -.			
DR	MEMDEL; J1569; Arath.Ypr9.J1569.			
DR	PROSITE; PS00435; PEROXIDASE_1; 1.			
DR	PROSITE; PS00436; PEROXIDASE_2; 1.			
DR	PFAM; PF00141; peroxidase; 1.			
DR	PRINTS; PR00458; PEROXIDASE.			
DR	PRINTS; PR00461; PLPEROXIDASE.			
KM	Peroxidase.			
SEQ	SEQUENCE	349 AA;	38108 MW;	0792EBF6 CRC32;
Query Match	67.1%; Score 1079.5;	DB 10;	Length 349;	
Best Local Similarity	68.2%; Pred. No. 2e-89;			
Matches 208;	Conservative 37;	Mismatches 59;	Indels 1;	Gaps 1;
Oy		2	OUTTFEYNDSGCNVSNIVADTVNELRSPPRIAAISILRHFDHFCVNGCDASILLDNSTS	61
Db		30	QLRPDPVFCTCFVEDIIGNIIVDELQTDPRIAASILRLHFHDFVCRCDDASILLDNSTS	89
Oy		62	FETERDAFCGNANASAGFPYVIDRMKAAYESACPRTVSCADLTITAAOQSVTLAGGPSMRVP	121
Db		90	FETERDAAPNMANASAGFNIVIDRMKALEBAECGRSGCADILITIAISOISLSSGHTFGMWVPV	149
Oy		122	LGRDSLQAFDLNANANI.PAPEFTLLPOLKDSFRNYGLNRSSDLVALSGGHTFGKNOCRTI	181
Db		150	LGRDSVEAFALANTWALPSPSTLTQKTARFADVGILNNPSDLVALSGGHTFGKNOCQCV	209
Oy		182	MRLTYNFSTGTLPDTLTNTTYIQTIRGLCPLNGNISALVDPLRPTTIIDNKYYVNLREQ	241
Db		210	TPLTNFMGNTPDPSLNPTYLVLELRRLCPONGNGTIVLVNFDSVPTPEDROQYTNLLNG	269
Oy		242	KGLISDOELFSSPATPTPIVRBSFANSSTOFENFAPEAMORMNITPLOTGOIRIN	301
Db		270	KGLISDOVLSTPCA-DTIPLVNOYSSTMTVFCAFDAMIRMKNLAKPLTGTOGEIRON	328
Oy		302	CRVVN	306
Db		329	CRVVN	333

ID	NAME	DESCRIPTION	SEQUENCE
OY	242	KG10SDDELSSDARDTLP1VAFSFSNQTQFFENAVEMADMGNRTPLTGTQGGRLRN	301
Db	270	KG10SDDELSTEGC-DT1PLVNOYSSDVSFFRAFDAMIRGNLRPLTGTGTGELRON	328
OY	302	CRVN 306	
Db	329	CRVN 333	
RESULT	5		
Q43100			
ID	Q43100	PRELIMINARY;	PRT; 343 AA.
AC	Q43100:		
DT	01-NOV-1996	(TREMBLrel. 01, Created)	
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)	
DE	01-NOV-1999	(TREMBLrel. 12, Last annotation update)	
DE	PEROXIDASE (EC 1.11.1.7) (MYELOPEROXIDASE).		
GN	YPR9.		
OS	Populus trichocarpa (Western balsam poplar).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	euphylliphytes; Spermatophyta; Magnoliophyta; eudicotyledons;		
OC	cory eudicotis; Rosidae; eustosids I; Malpighiales; Salicaceae; Populus		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=XYLEM;		
RA	CHRISTENSEN J.H., BAUM G., BOERJAN W., VAN MONTAGU M.;		
RL	Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.		
CC	-1- CATALYTIC ACTIVITY: DONOR + H(2)O(2) = OXIDIZED DONOR + 2 H(2)O.		
CC	-1- COFACTOR: HEME.		
DR	EMBL: X97349; CAAB6035.1; -		
DR	HSSP: P00433; 2ATJ		
DR	MENDEL: 10232; Poplr; YPR9; 10232.		
DR	PROSITE: PS00435; PEROXIDASE_1; 1.		
DR	PROSITE: PS00436; PEROXIDASE_2; 1.		
DR	PFAM: PF00141; peroxidase; 1.		
DR	PRINTS: PR00458; PEROXIDASE.		
DR	PRINTS: PR00461; PLPEROXIDASE.		
KW	Peroxidase; Oxidoreductase.		
SO	SEQUENCE 343 AA; 36568 MW; BFBE2DB CRC32;		

Query Match	65.4%	Score 1052	DB 10	Length 343
Best Local Similarity	64.5%	Pred. No. 5.8e-87		
Matches 198	Conservative	43	Mismatches 66	Indels 0
			Gaps	0
QY 2 QLTPEFYDNCSCVNNVNIYADTIVNLRSPDRIRNAILRLHFHDFCVCNCGDASILLDNRTTS 61	:       :	:       :	:       :	:       :
Db 25 QLTPEFYDQTCNVSSIIIRDYITETFLVSDPRIGASLIIRHFFHDFCVCNCGDASILLDNRTT 84				
QY 62 FTETEDAGNANASAGCFYVIDRMAKAAVESACPRYVSCADLLTTIAQOSVTLAGGFSWRVP 121	:     :     :     :     :     :     :     :     :	:     :     :     :     :     :     :     :     :	:     :     :     :     :     :     :     :     :	:     :     :     :     :     :     :     :     :
Db 85 IYSEKEAGGNNSNASGFEVVDTKAKLLESACPAITVSCADILLIAEESVYLAGGPNMTVP 144				
QY 122 LGRRLSLAFLDLANANLPAPFETLPOLKDSFRNGLNRRSGFLVLAISGHTFGPKRQCRPI 181	:     :     :     :     :     :     :     :     :	:     :     :     :     :     :     :     :     :	:     :     :     :     :     :     :     :     :	:     :     :     :     :     :     :     :     :
Db 145 LGRRRSTRASRAANAALPAPFETLIDQLRESTVNSLNNNSDLVALSGAHTFGRAKCSVF 204				
QY 182 MDRLVNSFNTGLPDDTLNTTYLQTLRGCLPLGNLSALVDFDLRTPIFDNRYVYVNEEO 241	:     :     :     :     :     :     :     :     :	:     :     :     :     :     :     :     :     :	:     :     :     :     :     :     :     :     :	:     :     :     :     :     :     :     :     :
Db 205 DRLRLDFNSTGAPDPSLDTTLTLLAQELCEEGGNSVITDLDLSTPPDAFDSDYVSNLQCN 264				
QY 242 KELLISDDELFSPPATDIIPLVRSFANSTOFFENAFPEANDRMGNITPPLTGTQOQIRLN 301	:       :     :     :     :     :     :     :     :	:       :     :     :     :     :     :     :     :	:       :     :     :     :     :     :     :     :	:       :     :     :     :     :     :     :     :
Db 265 KELLTIDDELFSPPADVDIALVNAFSAQITAFESFVSMIRMGNTLSPLTGTETEIRLN 324				
QY 302 CRVNVSN 308	:			
Db 325 CSVYMAN 331				
RESULT 6				
Q040950				

ID	Q040950.	PRELIMINARY:	PRT:	343 AA.
DT	01-NOV-1996	(TREMBLrel. 01, Created)		
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)		
DT	01-NOV-1999	(TREMBLrel. 12, Last annotation update)		
DE	PEROXIDASE.			
GN	YPR9.			
OS	Populus nigra (Lombardy poplar).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;			
OC	core eudicots; Rosidae; eurosids I; Malpighiales; Salicaceae; Populus.			
DR	[1]			
DR	SEQUENCE FROM N. A.			
RA	SAKUMA Y., AZUMA T., KATO Y., KOJIMA Y., MURA K.;			
RT	"Poplar peroxidase gene responing to cytokinin treatment.";			
RL	Submitted (Jan-1996) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: D83225; BAA11553.1; -			
DR	HSSP: P00433; 2ART.			
DR	MENDEL: 12194; Popnu:YPR9.12194.			
DR	PROSITE: PS00435; PEROXIDASE_1; 1.			
DR	PROSITE: PS00436; PEROXIDASE_2; 1.			
DR	PFAM: PF00141; peroxidase; 1.			
DR	PRINTS: PR00458; PEROXIDASE.			
DR	PRINTS: PR00461; PLPEROXIDASE.			
KW	Peroxidase.			
SO	SEQUENCE 343 AA; 36532 MW; 1645A1A5 CRC32;			
Query Match	65.0%; Score 1046; DB 10; Length 343;			
Best Local Similarity	63.2%; Pred. No. 2e-86;			
Matches 194; Conservative 47; Mismatches 66; Indels 0; Gaps 0;				
OY	2 QLTPTFYDNCSPVNSIVNVTIVNELRSDPRIMASILRLHFHDCVNGCDASILLDNNTTS 61			
DB	25 QLTPTFYDQTCPNVNSIIRNVTITELTSPDRIMASILRLHFHDCVNGCDASILLDNNTDT 84			
OY	62 FTETEDATGNNANSANGFPYIDMKAAVBSACRYTSCADLLTIAAQOSTYLAGGSMKRP 121			
DB	85 IESKEAAGNNNSANGFEVYDRMKALLBSACPATVSCADILLTIAEESVYLAGGPMITVP 144			
OY	122 LGRRSILOAFDLANANLPAPFFTLPOLKDSFRNNGLNKSPGLVALSGHFTGKNOCRPI 181			
DB	145 IGRROSTTASRAANAASLPAPFLPLDQLRESTNGLNNNTDVLVALSGHFTGKNOCRPI 204			
OY	162 MDRLVNSFTGLPDLTLNTTYLQTLRGICPLNGNLVALVDFDLRTPTTFDNKRYVNEEQ 241			
DB	205 DPLRLDFNSTGAPDPSLDPPTLLAALQELCPQCGNRSVITDIDLTPPDAFDSNYVNLGN 264			
OY	242 KGLIODELSSSPATDTITPLVBSFANSTGYFFNAFYEAEMDRMGNTITPLTGOCQIRLN 301			
DB	265 RGLLOTDELSTPGADVIAIVNAFSAHQTAFFESFAESMIRMGNLSPLTGTGEIEIRLN 324			
OY	302 CRVYNS 308			
DB	325 CRVYNS 331			
RESULT 7				
ID	Q040949	PRELIMINARY:	PRT:	343 AA.
AC	Q040949;			
DT	01-NOV-1996	(TREMBLrel. 01, Created)		
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)		
DT	01-NOV-1999	(TREMBLrel. 12, Last annotation update)		
DE	PEROXIDASE.			
GN	YPR9.			
OS	Populus nigra (Lombardy poplar).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;			
OC	core eudicots; Rosidae; eurosids I; Malpighiales; Salicaceae; Populus.			
DR	[1]			
DR	SEQUENCE FROM N. A.			
RA	SAKUMA Y., AZUMA T., KATO Y., KOJIMA Y., MURA K.;			

Query Match	64.9%	Score 1044	DB 10	Length 343
Best Local Similarity	63.8%	Pred. No. 3e-86		
Matches 196	Conservative 44	Mismatches 67	Indels 0	Gaps 0
RT	*Poplar peroxidase gene responding to cytokinin treatment.*			
RL	Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: D83224; BAA11852.1; -			
DR	HSSP: P00433; ZATJ.			
DR	MENDEL: 12193; Popul.Ypr9.12193.			
DR	PROSITE: PS00435; PEROXIDASE_1; 1.			
DR	PROSITE: PS00436; PEROXIDASE_2; 1.			
DR	PFAM: PF00141; peroxidase; 1.			
DR	PRINTS: PR00458; PEROXIDASE.			
DR	PRINTS: PR00461; PEROXIDASE.			
DR	PEROXIDASE.			
DR	SEQUENCE 343 AA; 36529 MW; ELAB207D CRC32;			
QY	2	QLTPEFYDSCNVSNIVKDTIVNELRSDPRIASILRLHFDCVYNGCDASILLDNFTS	61	
DB	25	QLTPEFYDSCNVSNIVKDTIVNELRSDPRIASILRLHFDCVYNGCDASILLDNFTS	84	
QY	62	FTEDEAFGNASAGFPVIDEMKAIVESACRPTYSACDLITLTAOOSYTLAAGPSWVRP	121	
DB	85	IVSEKAGGNNSAGFEVYDMKLLISACATYSCADILITLAAESVYLAAGPMTVP	144	
QY	122	IGRRSDIAFDLNLANLPAPEFTLPOLKDSFRNGLNRSSDVLVLSGGHFGKNOCFEI	181	
DB	145	IGRRSTTASRDAANAFLPAPEFTLDQLRESTVNSLNNNDIVLALSGHFGRAKCTF	204	
QY	182	MDRLNFSNTGLPDPYLTNTTYLQTLRGICPLNGNLISALVDFDLPTPTFDNRYVNEEQ	241	
DB	205	DFRLDFNFTSGAPDPSLNTTTLADLQELCPQGGNGSVYTDLDLTTPDAFDSNYYSNLGN	264	
QY	242	KLISDDELFSPPATPTIPLRSEFANSTOFFENAFVEMDRMKNINPLGTGSOQITLN	301	
DB	265	QGLLDITDDELSTPPADDIVIALVNAFSAHQTAFFESFAESMIRKGNLSPLTGTBETRLN	324	
QY	302	CRVYNS 308		
DB	325	CRVYNS 331		
RESULT	8			
ID	Q43101	PRELIMINARY:	PRT:	343 AA.
AC	Q43101:			
DT	01-NOV-1996 (Tremblrel. 01, Created)			
DT	01-NOV-1996 (Tremblrel. 01, Last sequence update)			
DT	01-NOV-1999 (Tremblrel. 12, Last annotation update)			
DE	PEROXIDASE (EC 1.11.1.7) (MYELOPEROXIDASE).			
GN	YPR9.			
OS	Populus trichocarpa (Western balsam poplar).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	euphylliphytes; Spermatophyta; Magnoliophyta; eudicotyledons;			
OC	core eudicots; Rosidae; eustosids I; Malpighiales; Salicaceae; Populus.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=XYLEM.			
RL	CHRISTENSEN J.H., BAUM G., BOERJAN W., VAN MONTAGU M.;			
RA	Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.			
CC	-1- CATALYTIC ACTIVITY: DONOR + H(2)O(2) -> OXIDIZED DONOR + 2 H(2)O.			
CC	-1- COFACTOR: HEME.			
DR	EMBL: X97350; CAA66036.1; -			
DR	HSSP: P00433; ZATJ.			
DR	MENDEL: 10233; Popul.Ypr9.10233.			
DR	PROSITE: PS00435; PEROXIDASE_1; 1.			
DR	PROSITE: PS00436; PEROXIDASE_2; 1.			
DR	PFAM: PF00141; peroxidase; 1.			
DR	PRINTS: PR00458; PEROXIDASE.			
DR	PRINTS: PR00461; PEROXIDASE.			
DR	PEROXIDASE.			
DR	SEQUENCE 343 AA; 36750 MW; 353EB6C5 CRC32;			

Query Match	64.7%	Score 1040:	DB 10:	Length 343:
Best Local Similarity	63.5%	Pred. No. 6.9e-86:		
Matches 195:	Conservative 45:	Mismatches 67:	Indels 0:	Gaps 0:
QY	2	QUTPFYDSCNVSNIIVRDTIVNELRSOPRFAASILRLHFHDCVYNGCDASILLDNNTTS	61	:      : : :         :      :      :       :
DB	25	QUTPFYDRTCNVSSIIISNVTTELVSDBTRGASILIRLHFHDCVYNGCDGSLLDNDT	84	:      :      :      :      :      :      :      :
QY	62	FTEEDAFGNMANSAGFPYIDRMKAIVESCBRTVSCADLLTIAAQQSVTLAAGPSWRY	121	:      :      :      :      :      :      :      :
DB	85	ISESEKANGNNNSARGFEVYDREKALLESACPTVSCADIIAIAEESVLAGPNNVT	144	:      :      :      :      :      :      :      :
QY	122	IGRRSLQAFDLNLANANLPAFFFTLPOLKDSRRVNGLNSSDLVALSGGHTFGKQCRFI	181	:      :      :      :      :      :      :      :
DB	145	IGRRSTASTASRDAAANAFPAFFFTLDOLRESFTNVSNNNSDLVALSGAHFTGRCSTF	204	:      :      :      :      :      :      :      :
QY	182	MDRLNFSNTGJDPDLNNTYLOTLRGICPLNGNISALVDPDLRPTLPDNKYVNLREQ	241	:      :      :      :      :      :      :      :
DB	205	VERLDENGTGAPDSTIDPFLLEALOKLCPENGNSVITLDVTTADAFDSKYSNLQCN	264	:      :      :      :      :      :      :      :
QY	242	KGLISDDELFSPPDPTDITPLRSEFANSTQTFEFAFEMADRMGNITPLTGQOIRLN	301	:      :      :      :      :      :      :      :
DB	265	RELQTDGELFSTPADDDIALVNFANQTFEFSFVBSMRMGNISPLTGTBEIRLN	324	:      :      :      :      :      :      :      :
QY	302	CRVYNSN 308		:
DB	325	CRVYNNAN 331		:
RESULT	9			
Q43099	Q43099	PRELIMINARY:	PRT:	343 AA.
AC	Q43099:			
DT	01-NOV-1996	(TREMblrel. 01, Created)		
DT	01-NOV-1996	(TREMblrel. 01, Last sequence update)		
DT	01-NOV-1999	(TREMblrel. 12, Last annotation update)		
DE	PEROXIDASE (EC 1.11.1.7) (MYELOPEROXIDASE).			
GN	YPR9.			
OS	Populus trichocarpa (Western balsam poplar).			
OC	Eukaryota: Viridiplantae: Streptophyta; Embryophyta; Tracheophyta;			
CC	euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;			
CC	corylidoites; Rosidae; eustosids I; Malpighiales; Salicaceae; Populus.			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=XYLEM;			
RA	CHRISTENSEN J H., BAUM G., BOERJAN W., VAN MONTAGU M.;			
RL	Submitted (Apr-1996) to the EMBL/Genbank/DBJ databases.			
CC	-1- CATALYTIC ACTIVITY: DONOR + H(2)O(2) - OXIDIZED DONOR + 2 H(2)O.			
CC	-1- COFACTOR: HEME.			
DR	EMBL: X97348; CA66034.1; -.			
DR	HSSP: P00433; 2ART.			
DR	MENDEL: 10231; Poptr; YPR9.10231.			
DR	PROSITE: PS00435; PEROXIDASE_1; 1.			
DR	PROSITE: PS00436; PEROXIDASE_2; 1.			
DR	PFAM: PF00141; peroxidase; 1.			
DR	PRINTS: PR00458; PEROXIDASE.			
DR	PRINTS: PR00461; ELPEROXIDASE.			
KW	Peroxiase; Oxidoreductase.			
SQ	SEQUENCE 343 AA: 36576 MW; DCEBD28D CRC32;			
Query Match	63.5%	Score 1028:	DB 10:	Length 343:
Best Local Similarity	62.5%	Pred. No. 8.4e-85:		
Matches 192:	Conservative 45:	Mismatches 70:	Indels 0:	Gaps 0:
QY	2	QUTPFYDSCNVSNIIVRDTIVNELRSOPRFAASILRLHFHDCVYNGCDASILLDNNTTS	61	:      : : :         :      :      :       :
DB	25	QUTPFYDRTCNVSSIIIRNVTTELVCDBCRIRGSLIRLHFHDCVYNGCDGSLLDNDT	84	:      :      :      :      :      :      :      :
QY	62	FTEEDAFGNMANSAGFPYIDRMKAIVESCBRTVSCADLLTIAAQQSVTLAAGPSWRY	121	:      :      :      :      :      :      :      :
DB	85	ISESEKANGNNNSARGFEVYDREKALLESACPTVSCADIIAIAEESVLAGPNNVT	144	:      :      :      :      :      :      :      :
QY	122	IGRRSLQAFDLNLANANLPAFFFTLPOLKDSRRVNGLNSSDLVALSGGHTFGKQCRFI	181	:      :      :      :      :      :      :      :
DB	145	IGRRSTASTASRDAAANAFPAFFFTLDOLRESFTNVSNNNSDLVALSGAHFTGRCSTF	204	:      :      :      :      :      :      :      :
QY	182	MDRLNFSNTGJDPDLNNTYLOTLRGICPLNGNISALVDPDLRPTLPDNKYVNLREQ	241	:      :      :      :      :      :      :      :
DB	205	VERLDENGTGAPDSTIDPFLLEALOKLCPENGNSVITLDVTTADAFDSKYSNLQCN	264	:      :      :      :      :      :      :      :
QY	242	KGLISDDELFSPPDPTDITPLRSEFANSTQTFEFAFEMADRMGNITPLTGQOIRLN	301	:      :      :      :      :      :      :      :
DB	265	RELQTDGELFSTPADDDIALVNFANQTFEFSFVBSMRMGNISPLTGTBEIRLN	324	:      :      :      :      :      :      :      :
QY	302	CRVYNSN 308		:
DB	325	CRVYNNAN 331		:

Dd	85	I ESEKEAGNNNSANGFEVYDMMKLLLESACATVCADILITIAAEESVYLAAGCMMTVP	144
Oy	122	LGRDSLOAEFLDLANANI.PAPEFTTLPOLKDFERNGLNRSSDYALSGGHTFGKNOCRFI	181
Dd	145	LGRBSSTASRAANAASIPAPFLTIDOLREFTNGNLMNTDVLVLSGAHFGRAKCSTF	204
Oy	182	MRLYNFSMTGLPDPTLWNTTYLOTIRGLCPLNGNL.SALVDPLRPTTFDKNKYYNLEEQ	241
Dd	205	NERLDVFNETGAPDPTLPOLPEFLAAQOEICPOGGNDVSVTDDLTTPDAEDSNYSMLCCN	264
Oy	242	KGLISDOELFESSPATDTPIPLVRFSNSTOTFFNAFVEAMDRMGNIPLTGTGOGIRLN	301
Dd	265	RCLLTGTDELEFETPCADDVIALVNAFSANQTRFFESFPESHMRGMNLSPLTGTEBEIRLN	324
Oy	302	CHRVANSN 308	
Dd	325	CNVNAN 331	
RESULT	10		
ID	Q43050	PRELIMINARY:	PRT, 343 AA.
AC	Q43050:		
DT	01-NOV-1996 (TREMBLrel. 01, Created)		
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)		
DE	PEROXIDASE (EC 1.11.1.7).		
DP	PEROXIDASE (EC 1.11.1.7).		
GN	PRXAAA.		
OS	Populus kitchakmensis (Aspen).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;		
OC	cory endicots; Rosidae; eurosids I; Malpighiales; Salicaceae; Populus,		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE: 95375232.		
RA	OSAKABE K., KOYAMA H., KAWAI S., KATAYAMA Y., MOROHOSHI N.;		
RT	"Molecular cloning of two tandemly arranged peroxidase genes from		
RT	Populus kitchakmensis and their differential regulation in the stem."		
RL	Plant Mol. Biol. 28:677-689(1995).		
DR	EMBL: D38051; BAA07241.1; -		
DR	HSSP: P00433; ZATP.		
DR	MENDEL: 12185; Popk1; ypr9.12185.		
DR	PROSITE: PS00435; PEROXIDASE_1;		
DR	PROSITE: PS00436; PEROXIDASE_2; 1.		
DR	PFAM: PF00141; peroxidase; 1.		
DR	PRINTS: PR00458; PEROXIDASE.		
DR	PRINTS: PR00461; PLPEROXIDASE.		
KW	peroxidase; oxidoreductase.		
FT	CHAIN 25 343		
SQ	SEQUENCE 343 AA: 37220 MM; 83FC34DI CRC32;		
Query Match	62.3%;	Score 1002;	DB 10; Length 343;
Best Local Similarity	62.2%;	Pred. No. 1.8e-82;	
Matches 191; Conservative	38;	Mismatches 78;	Indels 0; Gaps 0;
Oy	2	QLTPTFYDNCSGVNSIVYDTIVNELSRSPRIAASTLRHLHFDCVNGGDASILLDNPTS	61
Dd	25	QLTPTFYDETCTCYVIISTINGVIAETLIIDPRKIGASILRIHLHFDCVNGCDGSILLDKAT	84
Oy	62	FTEKDAPCFNANSARKGPVYIDRMKAVERESACPRTVSCADLITIAAQSVTLAAGSWRP	121
Dd	85	IDTEKEVALNANNSSARGFDVVDMKERLEGCVPAVSCADIILAAIEESVYLAVGPMWP	144
Oy	122	LGRRSLOAEFLDLANANLPAPFTTLPOLKDSRRNGLNRSSDYALSGGHTFGKNOCRFI	181
Dd	145	LGRROSLTANRRTAANAFTIGPQDTERLERSRTYVVGGLNMNTDVLVLSGAHFGRAQC	204
Oy	182	MRLYNFSMTGLPDPTLWNTTYLOTIRGLCPLNGNL.SALVDPLRPTTFDKNKYYNLEEQ	241
Dd	205	IBRLINENNTGTPDPTLDTTYLATIQRCLCPGGKNTIVLADLDPPTPGDFDNNYFSSNDAS	264
Oy	242	KGLISDOELFESSPATDTPIPLVRFSNSTOTFFNAFVEAMDRMGNIPLTGTGOGIRLN	301

```
DB 265 KGLQSDQELFSTPEADITELVDITFSTDETAFFESFVESMIRGNLSPLTGTGETRLN 324
QY 302 CRYVNSN 308
DB 325 CRAVNAD 331

RESULT 11
Q43051
ID 043051 PRELIMINARY: PRT: 314 AA.
AC 043051:
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-JAN-1998 (Tremblrel. 05, last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, last annotation update)
DE PEROXIDASE (EC 1.11.1.7) (MYELOPEROXIDASE) (FRAGMENT).
GN YPR9.
OS Populus kitakamiensis (Aspen).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids I; Malpighiales; Salicaceae; Populus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=STEM.
RA OSAKABE K., KOYAMA H., KAWAI S., KATAYAMA Y., MOROHOSHI N.;
RL Plant Sci. 103:167-175(1994).
CC -1- CATALYTIC ACTIVITY: DONOR + H(2)O(2) = OXIDIZED DONOR + 2 H(2)O.
CC -1- COFACTOR: HEME.
DR EMBL: D30652; BAA06334.1; -.
DR HSSP: P00433; 2ATU.
DR MENDEL: 12186; Poptr.Ypr9;12186.
DR PROSITE: PS00435; PEROXIDASE_1; 1.
DR PFAM: PF00141; peroxidase; 1.
DR PFM: PF00141; peroxidase; 1.
KW Peroxidase; Oxidoreductase.
FT NON_TER 1
SQ SEQUENCE 314 AA; 33471 MW; 9501F77A CRC32;

Query Match 61.9%; Score 996; DB 10; Length 314;
Best Local Similarity 62.6%; Pred. No. 5.6e-82;
Matches 189; Conservative 41; Mismatches 72; Indels 0; Gaps 0;

QY 7 FYDNCSPVNSIVRDTIVNELRSDPRIAASILRLHFHDCFVNGCDASILLDNSTFTEK 66
DB 1 FYDTCPNVSTIIRVDITETFLASPRIGASILRLHFHDCFVNGCDGSLLDNSDTIYSEK 60

QY 67 DARGNANSARGFVIDRKAAYESACPRVSCADLITTAQOQSYTLACGSMWRVPLGRD 126
DB 61 EAGGNNSARGFEVDKMLLESACPRVSCADLITTAEESEVLAGGPMWVPLGRD 120

QY 127 SLQAFDLANANPAPFETLPOLKDSFRNVLKRRSDVLVALSGHFGKNGCFRIMRLY 186
DB 121 STASRDAAANAFIPAPRITITDOLRESTVNGLNNSDVLVALSGAHTFGRAKCSFPRRLY 180

QY 187 NESNTGLPDEPLTNTYVLTQRLGCLPLNGNLISALVDEDLRTPTIEDNKYYVLEQKLIQ 246
DB 181 DFNSTGAPDQSLDPTLLAALQELCPQGGNSVLTLDLTPDAFDSNYYSNLQNGOQLIQ 240

QY 247 SDELTFSSPADPTITPLVRSFANSTOTFFNAFVEMADRMGNITPLVGTQOQIRLNCVYN 306
DB 241 TIDVLTSTPCADIVIALVNAFSANQTAFFESFASMSIRMGNLRLPLGTGTEGEIRLNCVYN 300

QY 307 SN 308
DB 301 AN 302

RESULT 12
Q43102 PRELIMINARY: PRT: 345 AA.
AC 043102:
DT 01-NOV-1996 (Tremblrel. 01, Created)
```

```
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
DE 01-NOV-1999 (Tremblrel. 12, last annotation update)
DE PEROXIDASE (EC 1.11.1.7) (MYELOPEROXIDASE).
GN YPR9.
OS Populus trichocarpa (Western balsam poplar).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids I; Malpighiales; Salicaceae; Populus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=XYLEM.
RA CHRISTENSEN J.H., BAUM G., BOERJAN W., VAN MONTAGU M.;
RL Submitted (APR-1996) to the EMBL/Genbank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: DONOR + H(2)O(2) = OXIDIZED DONOR + 2 H(2)O.
CC -1- COFACTOR: HEME.
DR EMBL: X97351; CAA66037.1; -.
DR HSSP: P00433; 1ATU.
DR MENDEL: 10234; Poptr.Ypr9;10234.
DR PROSITE: PS00435; PEROXIDASE_1; 1.
DR PROSITE: PS00436; PEROXIDASE_2; 1.
DR PFAM: PF00141; peroxidase; 1.
DR PRINTS: PR00458; PEROXIDASE.
DR PRINTS: PR00461; PEROXIDASE.
KW Peroxidase; Oxidoreductase.
SQ SEQUENCE 345 AA; 36648 MW; AFA48C40 CRC32;

Query Match 61.9%; Score 995.5; DB 10; Length 345;
Best Local Similarity 62.0%; Pred. No. 7.1e-82;
Matches 191; Conservative 43; Mismatches 73; Indels 1; Gaps 1;

QY 2 QLPTEFYDNCSPVNSIVRDTIVNELRSDPRIAASILRLHFHDCFVNGCDASILLDNSTTS 61
DB 29 QLPTEFYDGTCPNVSTIRGVLAQALQTPDRIGASILRLHFHDCFPDGGCGSLLDNITDI 88

QY 62 FRTFKDAFGNANSARGFVIDRKAAYESACPRVSCADLITTAQOQSYTLACGSMWRVPL 121
DB 89 IESEKEAPNNNSARGFVDVDMKAAYENACPGIVSCADLITTAEESEVLAGGPMWVPL 148

QY 122 LGRDLSQAFDLANANPAPFETLPOLKDSFRNVLKRRSDVLVALSGHFGKNGCFRIMRLY 181
DB 149 LGRDLSITANRSCANSSTIPASSELAVLWSKFAVAGLNTSSDVLVALSGAHTFGRAOCLNF 208

QY 182 MDRLYNFSNTGLPDEPLTNTYVLTQRLGCLPLNGNLISALVDFDLRTPTIEDNKYYVLEEQ 241
DB 209 ISRLYNFSGSGNDPPLTNTYVLTQRLGCLPLNGNLISALVDFDLRTPTIEDNFSNLQTN 268

QY 242 KGLQSDQELFSSPDAITITPLVRSFANSTOTFFNAFVEMADRMGNITPLVGTQOQIRL 301
DB 269 EGLQSDQELFSTTGA-DTIAIVNFSNOTAFFESFVSMIRGNLSPLTGTGETRLN 327

QY 302 CRYVNSN 309
DB 328 CRIVNSN 335

RESULT 13
Q43049 PRELIMINARY: PRT: 347 AA.
AC 043049:
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, last annotation update)
DE PEROXIDASE PRECURSOR (EC 1.11.1.7) (PEROXIDASE) (MYELOPEROXIDASE).
GN YPR9 OR PRX43A.
OS Populus kitakamiensis (Aspen).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids I; Malpighiales; Salicaceae; Populus.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE: 95375232.
RA OSAKABE K., KOYAMA H., KAWAI S., KATAYAMA Y., MOROHOSHI N.;
```

RT "Molecular cloning of two tandemly arranged peroxidase genes from  
 RL Populus kirkamienensis and their differential regulation in the stem."  
 RL Plant Mol. Biol. 28:677-689(1995).  
 RN [2]  
 RP SEQUENCE OF 40-347 FROM N.A.  
 RC TISSUE=STEM;  
 RA OSKABE K., KOYAMA H., KAWAI S., KATAYAMA Y., MOROHOSHI N.;  
 RL Plant Sci. 103:167-175(1994).  
 CC -1- CATALYTIC ACTIVITY: DONOR + H(2)O(2) = OXIDIZED DONOR + 2 H(2)O.  
 CC -1- COFACTOR: HEME.  
 DR EMBL: D38050; BAA07240.1; -;  
 DR HSSP: D30653; BAA06335.1; -;  
 DR HSSP: P00433; ZATU.  
 DR MENDEL: 12184; Popk1.Ypr9;12184.  
 DR PFAM: PF00141; peroxidase; 1.  
 DR PRINTS: PR00458; PEROXIDASE.  
 DR PRINTS: PR00461; PLPEROXIDASE.  
 KW Oxidoreductase; Peroxidase.  
 FT CHAIN 30 347  
 SQ SEQUENCE 347 AA; 37079 MW; 30D3FE8B CRC32;

Query Match 60.7%; Score 975.5; DB 10; Length 347;  
 Best Local Similarity 62.0%; Pred. No. 4.5e-80;  
 Matches 191; Conservative 39; Mismatches 77; Indels 1; Gaps 1;

QY 2 QLTPTFYDNCSPVNSVIRDTIVNELRSDPRIASILRLHFHDCFVNGCDASILLDTTTS 61  
 Db 30 QLTPTFYDNCSPVNSVIRDTIVNELRSDPRIASILRLHFHDCFVNGCDASILLDTTDT 89  
 QY 62 FRTKEKAFGNANSARGFVIDRKAAYESACPTVSCADLLITAAQSVTLAGSPSWRP 121  
 Db 90 ISEKEEAPNNNSVGRGDVDDMKALENACPGIVSCADLLITAAEQSVTLAGSPSWTVP 149  
 QY 122 LGRDLSIAQLFLANANLPAFFTLPOLKDSFRVNGLNSSDLVALSGGTFEGKNCRT 181  
 Db 150 LGRDLSIARSGANSALPSPFASLVYLSKFAVGLDTSDDLVALSGAHTFGRAQCSSF 209  
 QY 182 MDRLYNSNTGLDPPPTNTYTLQTLRGICPLNGLSALVDFDLRTPIFPNKYYVNEQ 241  
 Db 210 NLRLYFSGSGNDPPLNTTYTLAELOQLCPQAGNESVYTLMDPTTPTFGNTFSNLQTN 269  
 QY 242 KGLIQDQELFSSPDATDITPLVRSFANSQTQTFENAFVEMADRMGNITPLTGTGQIRLN 301  
 Db 270 EGLLRSDQELFSTTGA-DTIDIVNFFSSNQTAFFESFVSMIRGNISPLTGTGDEIRLN 328

QY 302 CRVYNSNS 309  
 Db 329 CRRVNDNS 336

## RESULT 14

ID 042905 PRELIMINARY; PRT; 359 AA.  
 AC 042905;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, last sequence update)  
 DE 01-NOV-1999 (Tremblrel. 12, last annotation update)  
 DE PEROXIDASE PRECURSOR (EC 1.11.1.7) (MYELOPEROXIDASE).  
 OS Linum usitatissimum (Flax) (Linseed).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 OC core eudicots; Rosidae; eurosids I; Malpighiales; Linaceae; Linum.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-STORMONT CIRRUS; TISSUE=LEAF, AND STEM;  
 RA OMANN F., TYSON H.;  
 RL Aust. J. Plant Physiol. 23:773-789(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-STORMONT CIRRUS; TISSUE=LEAF, AND STEM;  
 RA OMANN F., TYSON H.;

RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: DONOR + H(2)O(2) = OXIDIZED DONOR + 2 H(2)O.  
 CC -1- COFACTOR: HEME.  
 DR EMBL: L07554; ABA7602.1; -;  
 DR HSSP: P00433; LATU.  
 DR MENDEL: 11888; Linus.Ypr9;11888.  
 DR PFAM: PF00141; peroxidase; 1.  
 DR PRINTS: PR00458; PEROXIDASE.  
 DR PRINTS: PR00461; PLPEROXIDASE.  
 KW Signal; Peroxidase; Oxidoreductase.  
 FT SIGNAL 1 22  
 FT CHAIN 28 359  
 SQ SEQUENCE 359 AA; 38197 MW; F8479674 CRC32;

Query Match 59.4%; Score 955.5; DB 10; Length 359;  
 Best Local Similarity 59.1%; Pred. No. 3e-78;  
 Matches 182; Conservative 45; Mismatches 80; Indels 1; Gaps 1;

QY 2 QLTPTFYDNCSPVNSVIRDTIVNELRSDPRIASILRLHFHDCFVNGCDASILLDTTTS 61  
 Db 28 QLTPTFYDNCSPVNSVIRDTIVNELRSDPRIASILRLHFHDCFVNGCDASILLDSAT 87  
 QY 62 FRTKEKAFGNANSARGFVIDRKAAYESACPTVSCADLLITAAQSVTLAGSPSWRP 121  
 Db 88 ILSEKQALGNNSVGRGDVDDMKQVEACPGVSCADLLITAAESVTLAGSPSWAVP 147  
 QY 122 LGRDLSIAQLFLANANLPAFFTLPOLKDSFRVNGLNSSDLVALSGGTFEGKNCRT 181  
 Db 148 LGRDLSIARSGANSALPSPFASLVYLSKFAVGLDTSDDLVALSGAHTFGRAQCSSF 207  
 QY 182 MDRLYNSNTGLDPPPTNTYTLQTLRGICPLNGLSALVDFDLRTPIFPNKYYVNEQ 241  
 Db 208 VGRLYNFSNTGCGDPTTINATFLETROIQCPONGNSVLTMDRTADAFNSNFTNLQTR 267  
 QY 242 KGLIQDQELFSSPDATDITPLVRSFANSQTQTFENAFVEMADRMGNITPLTGTGQIRLN 301  
 Db 268 EGLLRSDQELFSTP-GSDTIELVRFANQTAFFQSVNSMIRGNIPPPGSPSEIRN 326

QY 302 CRVYNSNS 309  
 Db 327 CRRVNSAS 334

## RESULT 15

ID 09XFL3 PRELIMINARY; PRT; 341 AA.  
 AC 09XFL3;  
 DT 01-NOV-1999 (Tremblrel. 12, Created)  
 DT 01-NOV-1999 (Tremblrel. 12, last sequence update)  
 DE 01-NOV-1999 (Tremblrel. 12, last annotation update)  
 DE PEROXIDASE 1 PRECURSOR (FRAGMENT).  
 GN FBPI.  
 OS Phaseolus vulgaris (Kidney bean) (French bean).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 OC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. MONCALM.  
 RA BLEE K.A., JUPE S.C., RICHARD G., BOWWELL G.P.;  
 RT "Molecular cloning and differential expression of six members of the  
 peroxidase family in French bean."  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF149277; AAD37427.1; -;  
 KW Peroxidase.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 341 AA; 36626 MW; C147F3A0 CRC32;

Query Match 57.5%; Score 924; DB 10; Length 341;  
 Best Local Similarity 59.4%; Pred. No. 1.9e-75;



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 12:36:40 ; Search time 73.09 Seconds  
(without alignments)  
350.531 Million cell updates/sec

Title: US-09-246-451-11

Perfect score: 2180  
Sequence: 1 TTTETIOSNANLAPLPHPVPE.....IVSGVQALPLVWDPAFTKAV 414

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 178050 seqs, 61884766 residues

Total number of hits satisfying chosen parameters: 178050

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : PIR\_64:\*\*  
1: pir1:\*\*  
2: pir2:\*\*  
3: pir3:\*\*  
4: pir4:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2176	99.8	415	1	O4PSCP campor 5-monooxyg
2	377	17.3	393	1	C70929 cytochrome P450 Rv
3	372	17.1	410	2	S39924 cytochrome P450meg
4	366	16.8	396	1	B69851 cytochrome P450 yj
5	364	16.7	397	1	S51594 cytochrome P450 my
6	360	16.5	398	1	H70807 cytochrome P450 Rv
7	346.5	15.9	404	2	T30231 cytochrome P450 -
8	340	15.6	403	2	B35401 cytochrome P450 10
9	338.5	15.5	428	1	F70729 cytochrome P450 Rv
10	335.5	15.4	410	1	O4B56M cytochrome P450 10
11	333.5	15.3	405	1	B42606 cytochrome P450 CV
12	333	15.3	410	1	E69611 cytochrome P450 cy
13	330	15.1	406	2	T35401 cytochrome P450 10
14	329	15.1	411	1	JC5859 polypeptide synthas
15	326	15.0	406	1	A48495 linalool 8-monooxy
16	324.5	14.9	376	1	G69679 polypeptide hydroxy
17	319.5	14.7	405	1	H70752 cytochrome P450 Rv
18	318	14.6	310	2	T44767 cytochrome P450 1i
19	318	14.6	428	1	A42971 cytochrome P450ter
20	316.5	14.5	385	1	G69594 cytochrome P450 bl
21	316.5	14.5	399	1	A53578 cytochrome P450 -
22	316	14.5	412	2	S24750 cytochrome P450 (s
23	313.5	14.4	410	2	JC4287 cytochrome P450 sc
24	304	13.9	438	1	E70515 cytochrome P450 Rv
25	303	13.9	402	2	JC5151 nitric-oxide reduc
26	303	13.9	404	1	JC5150 nitric-oxide reduc
27	303	13.9	417	1	S49051 cytochrome P450 ly
28	296.5	13.6	408	2	PD0007 cytochrome P450 no
29	296.5	13.6	408	2	JC5674 cytochrome P450 no

30	294.5	13.5	337	2	T30601 cytochrome P450 hy
31	293.5	13.5	406	1	S18531 cytochrome P450 er
32	291.5	13.4	433	1	B70677 cytochrome P450 Rv
33	287	13.2	381	1	S15809 cytochrome P450 cy
34	277.5	12.7	310	2	T44857 probable hydroxyla
35	266	12.2	412	1	B40634 erythromycin monoo
36	265	12.2	414	1	E70708 cytochrome P450 Rv
37	262.5	12.0	411	2	T36526 probable cytochrom
38	261.5	12.0	386	2	T30235 cytochrome P450 -
39	255	11.7	406	3	T17487 cytochrome P450 hy
40	253	11.6	401	1	I40208 cytochrome P450 Bu
41	250	11.5	411	1	JC4003 cytochrome P450 -
42	239	11.0	433	2	T44587 cytochrome P450 ho
43	236	10.8	402	1	A70707 cytochrome P450 Rv
44	234	10.7	396	1	H70730 cytochrome P450 Rv
45	226	10.4	391	3	T17486 cytochrome P450 hy

## ALIGNMENTS

RESULT 1  
O4PSCP  
campor 5-monooxygenase (EC 1.14.15.1) cytochrome P450 101 - Pseudomonas putida plasm  
N:Alternate names: cytochrome P450-CAM  
C:Species: Pseudomonas putida  
C>Date: 30-Apr-1982 #sequence,revision 31-Dec-1993 #text\_change 03-Mar-2000  
C:Accession: A25660; S34614; C60886; A00194  
R:Unger, B.P.; Gunsalus, I.C.; Sliagar, S.G.  
J. Biol. Chem. 261, 1158-1163, 1986  
A:Title: Nucleotide sequence of the Pseudomonas putida cytochrome P-450-cam gene and  
A:Reference number: A94678; MUID:66111751  
A:Accession: A25660  
A:Molecule type: DNA  
A:Residues: 1-415 <UNG>  
A:Cross-references: GB:M12546; NID:g151114; PIDN:AAA25760.1; PID:g151115  
R:Atamaki, H.; Koga, H.; Sagara, Y.; Hosoi, M.; Horituchi, T.  
Biochim. Biophys. Acta 1174, 91-94, 1993  
A:Title: Complete nucleotide sequence of the 5'-exo-hydroxycampor dehydrogenase gene  
A:Reference number: S34613; MUID:93326643  
A:Accession: S34614  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-42 <ARA>  
A:Experimental source: PGJ1, ATCC 17453; CAM plasmid  
R:Romeo, C.; Moriwaki, N.; Yasunobu, K.T.; Gunsalus, I.C.; Koga, H.  
J. Protein Chem. 6, 253-261, 1987  
A:Title: Identification of the coding region for the putidaredoxin reductase gene fro  
A:Reference number: A60886  
A:Accession: C60886  
A:Molecule type: DNA  
A:Residues: 408-415 <ROM>  
R:Haniu, M.; Armes, L.G.; Yasunobu, K.T.; Shastri, B.A.; Gunsalus, I.C.  
J. Biol. Chem. 257, 12664-12671, 1982  
A:Title: Amino acid sequence of the Pseudomonas putida cytochrome P-450. II. Cyanogen  
A:Reference number: A00194; MUID:33030788  
A:Accession: A00194  
A:Molecule type: protein  
A:Residues: 2-55,58-276, 'Q', 278-361, 'S', 363-407, 'N', 409-415 <HAN>  
C:Genetics:  
A:Gene: camC; CYP101  
A:Genome: Plasmid  
C:Function:  
A:Description: catalyzes hydroxylation of campor to yield 5'-exo-hydroxycampor; elec  
C:Superfamily: Pseudomonas plasmid campor 5-monooxygenase; cytochrome P450 homology  
C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygena  
F:246-380/domain: cytochrome P450 homology <CYP>  
F:358/Binding site: heme iron (Cys) (axial ligand) #status experimental

Query Match 99.8%; Score 2176; DB 1; Length 415;  
Best Local Similarity 99.8%; Pred. No. 5.1e-159;  
Matches 413; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

*(Continued)*

```
OY      1 TTETIOSNANLAPLPVPHVEHLVFEDFMTKPNPSNLASAGVOEAMVALQESNVPLDVTWRCNG    60  
          |||||  
Db       2 TTETIOSNAMIALPVPVPEHVLVEDFDVMYNPSNLASAGVOEAAMVALQESNPDLVWTRCNG    61  
  
OY      61 GHHIATRGOLIREAYEDRYHRHSSECPTPREAGEAYDTIPSSMDPEQGOFALAAOVVG        120  
          |||||  
Db       62 GHMIATARGQLIRREAYEDYRIHFSSCEPFPREAAGEAYDPITPSMDEPCOGFRALAQQVVG        121  
  
OY     121 MPVADKLERNIOELACSLIESLRPOGCNFTEDEYAEPPIRIFMLLAGLEPDIDPHLKYL        180  
          |||||  
Db     122 MPAVDXLERNIOELACSLIESLRPGOCNPFEDYAEPPIRFIMLLAGLEPDIDPHLKYL        181  
  
OY     181 TDQMTPDGSMFTAFAKEAKEYLDILPIITEORROKPGTDAISIVANGVNRPRTSDSEAKR        240  
          |||||  
Db     182 TDMTPRDGSMFTAEAKEALAYDYLIPIIEORROKPGTDAISIVANGVNCRPITSDEARK        241  
  
OY     241 MGCLLVGGIDTVNNFLSFSMEFLAKSPENHOELLIERPERIPAACEELLRRRSIVA DGR        300  
          |||||  
Db     242 MCGLLVGGIDTVNNFLSFSMEFLAKSPENHOELLIERPERIPAACEELLRRRSIVA DGH        301  
  
OY     301 LTSDFEFHGVALKKGDILLPOLMSGLDERKNACPMNHDFSOKSHHTFGCHSHLCLOQ         360  
          |||||  
Db     302 LTSDFEHGVQLKKGDQILLPOMLSGLDERENACPMNHDFSOKSHHTFGCHSHLCLOQ         361  
  
OY     361 HLAAREIIYTLKEMLTRIPDESIAPAQAIOHKSGIVSVQALLPVWDPATTTAV        414  
          |||||  
Db     362 HLARREIIYTLKEMLT RIPDSIAPGAIOHKSGIVSVQALLPVWD PATTTAV        415  
  
RESULT   2  
CY70929  cyclochrome P450 Rv1785c - Mycobacterium tuberculosis (strain H37RV)  
           N:Contains: oxidoreductase (EC 1.-.-.)  
           C:Species: Mycobacterium tuberculosis  
           C:date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999  
           C:Accession: C70929  
  
R:COLE, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,  
  ; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.  
  Rajandream, M.A.; Rogers, J.; Rutter, J.; Seeger, K.; Skelton, S.; Squares, S.  
  Nature 393, 537-544, 1998  
A:Authors :Seares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MID:98295987  
A:Accession: C70929  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-393 <COL>  
A:Cross-references: GB:A1022021; GB:AL123456; NID:g3250699; PIDN:CAAI7707.1; PID:e125459  
A:Experimental source: strain H37RV  
C:Genetics:  
C:Gene: Rv1785c  
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology  
C:Keywords: oxalodurectase
```

*(Continued)*

Db 194 MVAEERRRNPDPVLSQV---OIGEDPILSEIEVLGSHLLLAGDITVMAAGFSLLELAR 250  
OY 267 SPENHOELIERPERIPACCELLR-RFSLVADGRILFSDYEFGHVOLKKGOIILLPQMLS 325  
Db 251 RQQLRAMLRDNPCKQIRVETIEELVRLPEPSAPAPARTTPEPVVGKTLTPAGSPVRLCMNAV 310  
OY 326 GIDERRKNACPMHVDSEROKVSHTTFHGSHLCLGQHILARRIIVTLKEMLTRIPDFSIAP 385  
Db 311 NNDGSDAMSTDELVADGKVHHRMGCGGPHRLGSHLARLELTLLVGEMLNDIPDFELAP 370  
OY 386 --GAQIOHKSGIVSGVALPLWY 406  
Db 371 DYAPETIRPSKSFALKNPLPLRW 392

RESULT 3  
S39924  
cytochrome P450meg - Bacillus megaterium (ATCC 13368)  
C:Species: Bacillus megaterium  
A:Variety: ATCC 13368  
C:Date: 07-Oct-1994 #sequence-revision 23-Feb-1996 #text-change 04-Mar-2000  
C:Accession: S39924; S32216  
R:Rauschenbach, R.; Isernhagen, M.; Noeske-Jungblut, C.; Boldol, W.; Stewart, G.  
Mol. Gen. Genet. 241: 170-176, 1993  
A:Title: Cloning sequencing and expression of the gene for cytochrome P450meg, the st  
A:Reference number: S39924; MUID:94049677  
A:Accession: S39924  
A:Molecule type: DNA  
A:Residues: 1-410 <RAU>  
A:Cross-references: EMBL:Z21932; NTD:g288298; PIDN:CAN79985.1; PID:g288300  
A:Experiences: ATCC 13368  
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology  
C:Keywords: chromoprotein; heme; iron; metalloprotein  
C:355/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match	17.1%;	Score 372;	DB 2;	Length 410;
Best Local Similarity	30.1%;	Pred. No. 5.5e-21;		
Matches 109;	Conservative 53;	Mismatches 180;	Indels 20;	Gaps 7;
QY	63	WIATRGOLIREAYEDYRHFSSSE-----CPPIRAGEAYDFIP-TSMDDPEQROFRALA	115	
		: : : :   : : : :   : : : :   : : : :   : : : :		
Db	44	WNVFYEDVKKRVLSDYKHHFSSVKKRTTISVGDSEGSVPERIQTESDPPDHKKRSL	103	
		: : : :   : : : :   : : : :   : : : :   : : : :		
QY	116	NOVYAMPVVDKLENNIQELASLISLRPOGQCNTEDEAFPPIRITFMLAGLPEEDIP	175	
		: : : :   : : : :   : : : :   : : : :   : : : :		
Db	104	AAAFTPRSLOWNEPIQIADDELIGOMGQTEIDIVASLASPLPIIVADLMGVSPKORL	163	
		: : : :   : : : :   : : : :   : : : :   : : : :		
QY	176	HLKYLTDQMTFRPDGSGMTFAE-----AKEALYDLPIIRORROKPGDAISYANQV	228	
		: : : :   : : : :   : : : :   : : : :   : : : :		
Db	164	LEKKVVDLFLRFDREKREBEVDKLYQVAKETQYLYLTVQKRLNPADDIISDLKSEV	223	
		: : : :   : : : :   : : : :   : : : :   : : : :		
QY	229	NGRPITSDAARMGCLLVGLDIPVNFLSFSM-BFLAKSPREOELIERPERIPACEE	287	
		: : : :   : : : :   : : : :   : : : :   : : : :		
Db	224	DGEMFTDDEVVRTTMLIGAGVETTSHLANSFYSLYDXDKREVQELHENLDLVPQAVE	283	
		: : : :   : : : :   : : : :   : : : :   : : : :		
QY	288	LLR-FESLVAOGRILTSYEFERGVQKKGGDQILPQMLSGLDERKNACPMHVSER-QKV	345	
		: : : :   : : : :   : : : :   : : : :   : : : :		
Db	284	MLREFENLIKIDRYVKEENDLLGYELKGDSDSVVWMSAANDEMFEDPFTLNTHRPNNK	343	
		: : : :   : : : :   : : : :   : : : :   : : : :		
QY	346	SHTEGHGSHSLCGHLARRELIVLKEMLTRIPDFSIAPAQIOHK--SGIVSGVQAL	402	
		: : : :   : : : :   : : : :   : : : :   : : : :		
Db	344	KHLTPGNGPHFCGLAPLARLEAKTALTATFLKKFKHIEAVPSQLEPENLTDSATGQTLISL	403	
		: : : :   : : : :   : : : :   : : : :   : : : :		
QY	403	PL 404		
		: : : :		
Db	404	PL 405		
		: : : :		
RESULT	4			
E69851				
yttochrome P450 yJib - Bacillus subtilis				





Matches	107;	Conservative	69;	Mismatches	169;	Indels	34;	Gaps	8;
QY	55	WTRCN-----	GGHMITRGQILREAYEDYRHSS-----	ECFFITRAGEAYDIPT	101				
Dy	26	WMRAQPVFERDRNRGTAASVYQAVDAIROPELTFEAGGIRDPDAPL-----	MMI	76					
QY	102	SMDPEQGFOPALANQVVGMPVDKLENRIQELACSLIESLRPOGQCFTEEDYAEPPIR	161						
Dy	77	DMDDPAHLRRKLKVYAGTTRKKRVKDKIASIALCDTLIDAVACERECGFVRDLAPLEMA	136						
QY	162	IFMLLAGL-PEEDIPPHLK-----	YLTDOMTRPDGSMTFEAKAEALDYDILPIIEQRQ	213					
Dy	137	VIGDMIGVRPQRIQDMFLRMSDDLVTFLSSHVSQEDFQTTM-DAAFAAYNDPIRATIAARA	195						
QY	214	KRGDAISIVANGQVNGRPITISDEAKRRCGLLVGGDLTVVNFSLFSMEFLAKSEHNQE	273						
Dy	196	DPTDVLVAVSVSEVDGSRSLSDDELMTLILIGDETRHRLTSGTEQLLRNDQWDL	255						
QY	274	LIERPERIPACCELLRFRSLVAD-GRILTSDFEFGQLKGGDQILPOMUSGIDEKRN	332						
Dy	256	LQRPDSLLPGALIEELKRLRTAPVKKNRCRYLTADTEPHGALCAGEKMMILFESANDEAVF	315						
QY	333	ACPMDHVESROKVSHTTFGHSILCLGQHLARREIIVTLKMLTRIPDPSIAPGAQI--Q	390						
Dy	316	CEPEKFEVQQRPNDSHLAGFGTFHFCIGNQLARLELSTWTERVLRRLPDILRLVADDSVPL	375						
QY	391	HKSGIVSGVQALPLTWMDPA	409						
Dy	376	RPANVSGLESMPPVETPS	394						

RESULT 7  
T30231  
cytochrome P450 - Streptomyces hygroscopicus  
C:Species: Streptomyces hygroscopicus  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 31-Jan-2000  
C:Accession: J30231  
R:Apricito, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Staun-  
Gene 169, 9-16, 1996  
A:Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hygro-  
A:Reference number: Z20782  
A:Accession: T30231  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-404 <APA>  
A:Cross-references: EMBL:X86780; NID:9987088; PID:9987105; PIDN:CA60465.1  
C:Genetics:  
A:Gene: rapN  
A:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology

	Query Match	15.9%	Score 346.5	DB 2	Length 404
	Best Local Similarity	28.6%	Pred. No. 4.9e-19		
	Matches 102	Conservative 63	Mismatches 177	Indels 15	Gaps 9
QY	63	WIATRGOLIREAYEDYRHFSSCEPF--IPREAGEAYDF-IPTSMDPEQOFRALANOV	118		
		WLAASMEDVAVKVFWDPR-FSSRATLGKDVPRVLPAICQOPIYIMLDPPEHTRLRVRATKA	108		
Db	50				
QY	119	VGMVVDKLEKRIQELACSLIESLRPOC-QCNFEDYAEPPPIRATFMLLADLPEDDIPHL	177		
		LTSRMEELRPRTEVADDLIDKMLAKGAPADLMEDEFALPIPIIMICELLGVPIEDQYKF	168		
Db	109				
QY	178	KYLTDQMTRPDGSMT---FAEAKELAYDYLLPIIEORQRKGTDAISIVANGVNGRPRT	234		
		RTWSDQHL-SNGAASQEVMAAGOSLYIYISETLLAERKKOTNDLGLSVARAKDRDLS	227		
Db	169				
QY	235	SDEAKKRCGILLVGGDLTVNVFLSFSMEELAKSPHROELLERPRRIPACCELLRPSL	294		
		ETELVGFAYVLLINGVEYTTANAIGNSVYTLTHTEKLAELLRKDSLIRPAVDELLRIPI	287		
Db	228				
QY	295	VADG---RIILTSDEYFHFVQLKKGDOILLPQMLSGIDLRKNACCPMHVDFSQKVSHTTFG	351		

Db 268 AKQSNVRAVEDVELSGTLYKAGEAVAIOTHSANTDPKVIYDHEEITDFHRTSNPMKSLG 347

RESULT 8  
B35401  
cytochrome P450 105B1 - Streptomyces griseolus  
N:Contains: oxidoreductase (EC 1.-.-.-)  
C:Species: Streptomyces griseolus  
C:Date: 31-Mar-1992 #sequence.revision 31-Mar-1992 #text\_change 04-Mar-2000  
C:Accession: B35401  
R:Omer, C.A.; Lenthra, R.; Little, P.J.; Dean, C.; Tepperman, J.M.; Leto, K.J.; Rommess  
J. Bacteriol. 172, 3335-3345, 1990  
A:Title: Genes for two herbicide-inducible cytochromes P-450 from Streptomyces griseo  
A:Reference number: A35401, MUID:90264332  
A:Accession: B35401  
A:Molecule type: DNA  
A:Residues: 1-403 <OME>  
A:Cross-references: GB:M32239; NID:g153480; PIDN:AAA26825.1; PID:g153481; GB:M36481  
A:Genetics:  
A:Gene: CYP105B1  
A:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology  
C:Keywords: Chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygenase  
E:238-374/domain: cytochrome P450 homology <P45>  
F:352/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match	15.6%	Score 340;	DB 2;	Length 403;
Best Local Similarity	28.9%	Pred. No. 1.5e-18;		
Matches 105;	Conservative 63;	Mismatches 165;	Indels 30;	Gaps 13

OY		63	MIATRGOLIRAEVADYDHPHSESEC-----PIIPRAGEAYOFIPI--SMDDPEOROPRAA	115
Dd		52	WLAVTRHODVRAVLCDPR-FSDAHARTGPPFLUTACGREIICTNPFELRMDDPEHARLRML	110
OY		116	NOQVMGPVVDKLENRIOEIACSLIESLDP--OGOCNFTEDEYAEEPFIPIFMLGLGLEEDI	174
Dd		111	TADFIVAKVEAMREVOGRIMDDLVDRMTTGRTSADLTETEARPLPSLYICLLLGVEYEDH	170
OY		175	PHL---KYLTDOMTKPDGSKTFEAKREALYDLPIITEOROKPGTDAIS-IVANGVN	229
Dd		171	AFOERSKVILLTTLRSTPE---EVRAADEILEXYLARLAFKRERBDPAISRVLARGEID	227
OY		230	GRIPTSDCAKMGCLLVLVGGLDYVYNLFSSMELASPEHROELLERPRIPIACEBELL	289
Dd		228	DTQIAT-----MGRLILVAGHEITANNATLSTVLILRNPNOLARLRAEPALVGAVEELL	282
OY		290	RRESLVAVDG--RIITSYEFHGVOYLKGGDOIILPOMLSGDERRKNACP--MHYDFSRKOV	345
Dd		283	KYLTVIHNGVPRIATEDVYLIGGRITAAAGEVLC--MISSNNRADEVFPGGDDILDVARDAK	340
OY		346	SHTTEGHSHLICLOHLARREIIVTLKEWLTFRIPDFSIA-PGAOIHKSGI-VSGYOALP	403
Dd		341	RHVAFEGGVHQCLQGAPLAREVELQAIETLLRLRPDLRLAVPBHEIIPRGMALYGVHSLP	400
OY		404	LVM	406
Dd		401	IAW	403

RESULT 9  
F70729  
cytochrome P450 RV2266 - Mycobacterium tuberculosis (strain H37Rv)  
N:Contains: oxidoreductase (EC 1.-.-.-)  
C:Species: Mycobacterium tuberculosis  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: F70729  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garlier, T.; Churcher, C.; Harris, D.; Gordon  
Connort, R.; Davies, R.; Devlin, K.; Feldman, T.; Gentles, S.; Hamlin, N.; Holtroyd,  
A.; Indream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors : Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Accession number: A70500; MUID:98295987

A:Citation: F70729

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-428 <COL>

A:Cross-references: GB:277163; GB:A1123456; NID:g3261610; PIDN:CAB00969.1; PID:e255175;

A:Experimental source: strain H37Rv

C:Genetics:

C:Gene: RV2266

C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology

C:Keywords: oxidoreductase

F:264-401/Domain: cytochrome P450 homology <CYP>

Query Match 15.5%; Score 338.5; DB 1; Length 428;  
Best Local Similarity 26.2%; Pred.No.2.e-18;  
Matches 117; Conservative 72; Mismatches 188; Indels 69; Gaps 17;

OY 5 IOSNMNLAPLPHVPEHLVFDFDMYNPSMLSGVOGMAVLDE-----SNPDLV 54  
| | | | |  
DB 7 IATRVNGTP-PREVP---IADIEL-----GSLEFVALDDVDGDGAFATLRREARIPSE 54  
  
OY 55 WTR-----CMGGWIAIRGQILIREAYEDY----RH---FSS-----ECFFIPIREA 92  
| | | | |  
DB 55 WTIELPGFVNACGNHIALTK-----YDDEVYASHNPDIIFESSYPMITINDOPBELAEYF 107  
  
OY 93 GENAVIDFTSPDPREORGFALANOVGMPPVVDKLENRIQLACLSTESL--RPOGCCN 149  
| | | | |  
DB 108 GSM-----TVLDDPRHQRLRSIVSRAFPTRPVARIIEAAVRDHNRHVSSMIANNPROAD 162  
  
OY 150 FTEDIAEERPTIRFLMAGLPREDIPHLKYLVLDQMKT---RPDSMFPAEKALYLVD---Y 203  
  
OY 163 LVSELALGPPLPDICDMMGIFPKADHORIFHWNTNVILGFSDPDLDATFDDEFMOVSADI GA Y 222  
  
OY 204 LPIITEORROKRGDTAISIVANGVNGRRITSDEAKRMGLLLVGLDVTVNFSLFSME E 263  
| | | | |  
DB 223 ATVALEDARRVVNHHDLTSSLVEAEVDGERLSSEIASPFLLLVAACNETTNRNAITHGVLA 282  
  
OY 264 LAKSPEHRQELIERPERT-PAACEELLRFSS-LVADGRLLTSDEYFHGYOLKKGOAILP 321  
| | | | |  
DB 283 LSRYEPQRDRMWSDPDGLFAVEELVRNASPVVMRTLODIELRGTKMAAGDVSLSM 342  
  
OY 322 QMLSGUDEKKNMCMPMHVDSROKVSHTFE-GHGSHTCLGONLARREITYTLKEMLRIPD 380  
| | | | |  
DB 343 YCSANDESKEFKDPWFTEDLARNPNHLFGGGGAHFCLGANLARREIVAFADELROMPD 402  
  
OY 381 FSIAFGAQIQHKSGIVSQUALPLVW 406  
| | | | |  
DB 403 V-VATEEPARLLSQFIHGIKITLPTVM 427  
  
RESULT 10  
OABS6M  
cytochrome P450 106 - Bacillus megaterium  
N:Alternate names: cytochrome P450BM-1  
N:Contains: oxidoreductase (EC 1.-.-.-)  
N:Species: Bacillus megaterium  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 03-Mar-2000  
C:Accession: S07764; S17973  
R:He, J.S.; Ruettlinger, R.T.; Liu, H.M.; Fulco, A.J.  
Biochim. Biophys. Acta 1009, 301-303, 1989  
A:title: Molecular cloning, coding nucleotides and the deduced amino acid sequence of p  
A:Reference number: S07764; MUID:90089408  
A:Molecule type: DNA  
A:Residues: 1-410 <HEJ1>  
A:Cross-references: EMBL:X16610; NID:g339626; PIDN:CAA34612.1; PID:g339627  
A:Accession: S17973  
A:Molecule type: protein  
A:Residues: 1-25 <HEJ2>

C:Genetics:  
A:Gene: CYP106  
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homolog  
C:Keywords: Chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygenase  
F:1-410/Product: cytochrome P450 106 #status experimental <MAT>  
F:241-378/Domain: cytochrome P450 homolog <CYP>  
F:356/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 15.4%; Score 335.5; DB 1; Length 410;  
Best Local Similarity 27.7%; Pred. No. 3.4e-18;  
Matches 96; Conservative 57; Mismatches 170; Indels 23; Gaps 7;

QY 63 WATGQLIREAYEDYRHFSSP---CPPIREAGAYDIP---TSDPEQRALFA 115  
| : : : : | : : : : | : : : : | : : : : | : : : : |  
DB 45 WNVFYEHKQVLYSVDFSSDGORTTFVGNSKKSKSTPTNLTPDPHRRASLL 104  
  
QY 116 NQVGMPPVVDKLKNTLOELACLSIESLRQGCNTEDEYARFPRIIMLAGLPEEIDP 175  
| : : : : | : : : : | : : : : | : : : : | : : : : |  
DB 105 AAFTPRSLKNNEPKIQADLVLEAIQNSTINIVDDLSFFPSVLADLGVVVKORY 164  
  
QY 176 HLKYLTDMQTRPDGSMTFAEAKE-----ALDYLIPIEQRPKPCTDAISYANOV 228  
| : : : : | : : : : | : : : : | : : : : | : : : : |  
DB 165 OFKKWVDILFYODDERLEIEQEKORAGEFYQLYIYIEKRNLSDDIISDLQAEV 224  
  
QY 229 NGRPITSDAKRMCGILLVGLDTYVNF-L-SFSMEFLAKSPENROELIERPERIPACEE 287  
| : : : : | : : : : | : : : : | : : : : | : : : : |  
DB 225 DGEFTDEDEIVHATMLLAGAVETTSHAIANFYSLVDKSLYSSELNNRLAPKAVEE 284  
  
QY 288 LLR-FESLVADRILITSDYEFHGVLKGGDQILFQMISGLDERKNACPMHYDSR-QKV 345  
| : : : : | : : : : | : : : : | : : : : | : : : : |  
DB 285 MLRYRHHSIRSRDYIKODELLGVKLKGDDVIAMASCNDMETFENPFSDVHIRPTNK 344  
  
QY 346 SHTTFGHGSHLCIGOLARRRITVTLKEMLTR-----IPPSIAP 385  
| : : : : | : : : : | : : : : | : : : : | : : : : |  
DB 345 KILTGNGPHPCLGAPLARLEMKITLEAFLEAFSHIEPEDELEP 390

RESULT 11  
B42606  
cytochrome P450 CVIIIB - Saccharopolyspora erythraea  
N:Contains: oxidoreductase (EC 1.-.-.)  
C:Species: Saccharopolyspora erythraea  
C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 03-Mar-2000  
C:Accession: B42606  
R:Andersen, J.F.; Hutchinson, C.R.  
J. Bacteriol. 174, 725-735, 1992  
A>Title: Characterization of Saccharopolyspora erythraea cytochrome P-450 genes and e  
A:Reference number: A42606; MUID:92121109  
A:Accession: B42606  
A>Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA; protein  
A:Residues: 1-405 <AND>  
A:Cross-references: GB:M83110; NID:g152682; PIDN:AAA26483.1; PID:g152684  
A>Note: sequence extracted from NCBI backbone (NCBIP:77484)  
C:Genetics:  
A:Gene: CYP107B1  
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homolog  
C:Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase  
F:238-374/Domain: cytochrome P450 homolog <CYP>  
F:352/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 15.3%; Score 333.5; DB 1; Length 405;  
Best Local Similarity 25.6%; Pred. No. 4.8e-18;  
Matches 107; Conservative 74; Mismatches 184; Indels 53; Gaps 10;

QY 18 VPEHLVFD----FDMYNSNLSAGVOEAMAVLOESNVDPDLVTYRCNG-GHWIATRGOILR 72  
| : : : : | : : : : | : : : : | : : : : | : : : : |  
DB 6 VFDLIAFDAAFNQDKHN-----RYARKREPVRRI--RTYNGIDLAWLITYEDVK 53  
  
QY 73 EAVEDYRHFSSECPPIPREAGAYDIFPTSM-----DPEQRQFR 112

Db 54 QALLD-----PRIKDEFRTQOIIIEKRLADERRRPGSPDLGPHMLNTDPDHTRLR 105  
113 ALANOVGMVYDKLENRIOELACSLIESLRPOGOCNFTEDYAEPPIRIF 172  
106 KLVKATARRVEELRRPREDITDDLRAGSEVDIDEFAPPLPTTYSLEMGEDS 165  
173 DIPHLKYLTDQMTDPDGSMTFAEAK--ALYDYLPIIEORRQPGTDAISVANGVNG 230  
166 RRDFRSMTNVLV--DGSOPBAQAASVAMVEYLTELAKRTPEGDGLTALLEAVEDG 223  
231 RPTSDAKRRCGLLYGGLDVTYNFISFSEFLAKSPENHOLEIERPERIPACEELR 290  
224 DRLESEGLIMAVFLLVAGHETVNLGNCLSLGNPDQALARNPDSLPGAIIEETLR 283  
291 RFLSVAG--RLITSDFEFHGVOLKKGDDILLPQMLSGLDERKNACPMHWDFSKOVSH 348  
284 YESVYAKGTRHRTAEAVRFGDVVPEBGLVWALGAANRDEGEREDDPREDITRETTGHV 343  
349 TFGHSHLCLGQHLARREIIVTLKEWLRIPDFSIA--PGAQIOHKSQIVSGVALPL 404  
344 AFGCHIFCVCAALRLAEQIAGVRLLERPDLRMASPDLLMRFSVLMNGLEKLV 401

## RESULT 12

E69611

cytochrome P450 CYP4 - Bacillus subtilis

N:Contains: oxidoreductase (EC 1.-.-.-)

C:Species: Bacillus subtilis

C&gt;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 18-Feb-2000

C/Accession: E69611; T44774

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azavedo, V.; Bertet

C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd

A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Filtz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Gallier

lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.

Koelter, P.; Konungstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,

A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel

Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelli,

Rieger, M.; Rivolet, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,

A.; Authors: Schleich, S.; Schreuter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot

akech, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K

A.; Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A:Reference number: A69580; MUID:98044033

A:Accession: E69611

A&gt;Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-410 &lt;KUN&gt;

A:Cross-references: GB:299117; GB:AL009126; NID:g2634966; PIDN:CAB14615.1; PID:el183903;

A:Experimental source: strain 168

R:Belitsky, B.R.; Gustafson, M.C.U.; Sonenshein, A.L.; von Wachenfeldt, C.

J. Bacteriol. 179, 5448-5457, 1997

A:Title: An lrp-like gene of Bacillus subtilis involved in branched-chain amino acid tra

A:Reference number: 222837

A:Accession: T44774

A&gt;Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-410 &lt;BEL&gt;

A:Cross-references: EMBL:Y11043; PIDN:CAA71937.1

A:Experimental source: strain 1A1

C:Genetics: cypA

C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology

C:Keywords: oxidoreductase

F:245-381/Domain: cytochrome P450 homology &lt;CYP&gt;

QY 104 DPEQORFALANOVGMVYDKLENRIOELACSLIESLRPOGOCNFTEDYAEPPIRIF 163  
100 DPEQORFALANOVGMVYDKLENRIOELACSLIESLRPOGOCNFTEDYAEPPIRIF 159  
164 MLGAPDEEDIPHLKYLTDQMTDPDGSMTFAEAKAL-----YDYLPIIEORRQ 213  
160 SEMIGFLEDRKFRVWSQA-----IDFSAPERLOENDHLLGEFEVLEIVRRKR 213  
214 KPGTDAISVANGVNGVRPITSDAKRRCGLLYGGLDVTYNFISFSEFLAKSPENHOLEIERPERIPACEELR 273  
214 EPAGDLISALIQAESGETQSTEEYSIMILLVAGHETVNLGNCLSLGNPDQALARNPDSLPGAIIEETLR 273  
274 LIERPERIPACEELLRFSLV--ADGRITSDYEFHGVOLKKGDDILLPQMLSGLDERK 331  
274 LRQOPDLMAISIEALRPHSPVELTTRWAEPPILHGOELKRKDVIIISLASANREKI 333  
332 NACPMHWDFSRQKVSHTFGHSHLCLGQHLARREIIVTLKEWLRIPDFSIAPGAQIOH 391  
334 FPNADIDIERKNRHHAFGHNFCIGAQDLRLAKIAISTLLRCPNIQL-KGEKKOM 392  
QY 392 K---SGIVSGVALPL 404  
393 KWKGNFLMRALIELPI 408

## RESULT 13

A35401

cytochrome P450 105A1 - Streptomyces griseolus

N:Contains: oxidoreductase (EC 1.-.-.-)

C:Species: Streptomyces griseolus

C&gt;Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 04-Mar-2000

R:Omer, C.A.; Lenstra, R.; Little, P.J.; Dean, C.; Tepperman, J.M.; Leco, K.J.; Romess

J. Bacteriol. 172, 3335-3345, 1990

A:Title: Genes for two herbicide-inducible cytochromes P-450 from Streptomyces griseo

A:Reference number: A35401; MUID:90264332

A:Molecule type: DNA

A:Residues: 1-406 &lt;OME&gt;

A:Cross-references: GB:M32238; NID:g153477; PIDN:AAA26823.1; PID:g153478; GB:M36480

C:Genetics: cyp105A1

A:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology

C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygena

F:241-377/Domain: cytochrome P450 homology &lt;P45&gt;

F:355/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 15.1%; Score 330; DB 2; Length 406;  
Best Local Similarity 26.3%; Pred. No. 8.9e-18;

Matches 104; Conservative 74; Mismatches 193; Indels 24; Gaps 11;

QY 31 PSNLISAGV--EAMAVIQESNVDPDLWTRCNGH-WIATGQLIREAYEDYR----- 79  
17 PSNNSCPYQ--PDGAQADRPGLRHRTYLDGRQAVVYTHNEAARKLIGDPRSSNRTDD 76  
80 HFSSECFIP--REAGAYVFIPTSMOPREORFALANOVGMVYDKLENRIOELACS 137  
77 NFPAVSREFAVRESPAF---IGLDPREHGRTRRTTISEFTVKRKGRPREVEEVNG 132  
138 LIESLRPOG--QCNTEDYAEPPIRIFMLLAGPEEDIPHLKYLTDQMTDPDGSMTFAEA 196  
133 FLDEMLAAGPTADLVDSGFALPVSMTICRLGYPVADHFEFQDASKRVYSTQAQSLTA 192  
197 KEALYDLIPDIEORRQPGTDAI-SIVANGVNGRPTSDAKRRCGLLYGGLDVTYN 255  
193 RNDLAGYLDLITOFTEPEAGLVGALVADQLANGE-IDREELISTPMILLIGHETTA 251  
256 FLFSMEFLAKSPENHOLEIERPERIPACEELLRSL--VADGRITSDYEFHGVOLK 313  
252 MTSLSVITLLDHEQYAAALRADRSVLPGAVEELLRYLAADIAGRGVATDIEVGHILR 311





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2000, 13:04:51 ; Search time 27.02 Seconds  
(without alignments)  
474.815 Million cell updates/sec

Title: US-09-246-451-11  
Perfect score: 2180  
Sequence: 1 TTETIQSNANLAPLPPHVE.....IVSGVALPLVMDPATTKAV 414

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 85661 seqs, 30989116 residues  
Total number of hits satisfying chosen parameters: 85661

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_38.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2176	99.8	414	1	CPXA_PSEPU
2	591.5	27.1	414	1	Y4VG_RHISN
3	377	17.3	393	1	YH85_MYCTU
4	372	17.1	410	1	CPXM_BACME
5	366	16.8	396	1	YJIB_BACSU
6	361.5	16.6	405	1	CPXM_BACSU
7	360	16.5	398	1	Y218_MYCTU
8	340	15.6	402	1	CPXF_STRGO
9	338.5	15.5	428	1	YM66_MYCTU
10	335.5	15.4	410	1	CPXI_BACME
11	333.5	15.3	405	1	CPXK_SACER
12	333	15.3	410	1	CPXI_BACSU
13	330	15.1	405	1	CPXE_STRGO
14	319.5	14.7	405	1	YC56_MYCTU
15	318	14.6	428	1	CPXL_PSEPU
16	316.5	14.5	395	1	BIOL_BACSU
17	316.5	14.5	399	1	FAS1_RHOFA
18	316	14.5	412	1	CPXH_STRGR
19	304	13.9	438	1	Y180_MYCTU
20	303	13.9	408	1	NOR_FUSOX
21	296.5	13.6	402	1	NOR2_CYLTO
22	295	13.5	381	1	CPXG_STRSO
23	291.5	13.4	433	1	Y245_MYCTU
24	288.5	13.2	403	1	NORI_CYLTO
25	288.5	13.2	404	1	CPXJ_SACER
26	277.5	12.7	397	1	CPXO_SACER
27	265	12.2	400	1	CPXP_RHISN
28	265	12.2	414	1	P778_MYCTU
29	253	11.6	401	1	CPXP_BRAJA
30	236	10.8	402	1	P766_MYCTU
31	234	10.7	396	1	YM76_MYCTU
32	225	10.3	436	1	THCB_RHOER
33	220.5	10.1	400	1	YV21_MYCTU

34	217	10.0	422	1	CPXC_AGR6	P24466 agrobacteri
35	215	9.9	429	1	CPXR_BRAJA	O59204 bradyrhizob
36	206.5	9.5	489	1	YM68_MYCTU	O59572 mycobacteri
37	195	8.9	405	1	CYPX_BACSU	O34926 bacillus su
38	189	8.7	467	1	CPXR_RHISN	P55543 rhizobium s
39	177	8.1	368	1	CPXM_SULSO	O35080 sulfolobus
40	162	7.4	1048	1	CPXB_BACME	P14779 bacillus me
41	156.5	7.2	503	1	CP30_SHEEP	O29456 ovis aries
42	154.5	7.1	531	1	CP27_HUMAN	O02318 h cytochrom
43	151	6.9	407	1	CPXD_AGR6	P24467 agrobacteri
44	151	6.9	492	1	CPS3_PIG	O02390 sus scrofa
45	145.5	6.7	489	1	CPC3_RABIT	P00182 oryctolagus

## ALIGNMENTS

RESULT	1	CPXA_PSEPU	STANDARD:	PRT:	414 AA.
ID	CPXA_PSEPU	P00183:			
AC	21-JUL-1986 (Rel. 01, Created)				
DT	13-AUG-1987 (Rel. 05, Last sequence update)				
DT	15-FEB-2000 (Rel. 39, Last annotation update)				
DE	CYTROCHROME P450-CAM (EC 1.14.15.1) (CAMPHOR 5-MONOOXYGENASE).				
GN	CAMC OR CYP101.				
OS	Pseudomonas putida.				
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;				
CC	Pseudomonas.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE: 86111751.				
RA	Unger B.P., Gunsalus I.C., Silgar S.G.;				
RT	"Nucleotide sequence of the Pseudomonas putida cytochrome P-450cam				
RT	gene and its expression in Escherichia coli.";				
RL	J. Biol. Chem. 261:1158-1163(1986).				
RN	[2]				
RP	SEQUENCE OF 385-414 FROM N.A.				
RP	SPRAIN-ATCC 17453;				
PC	MEDLINE: 90130389.				
RX	Koga H., Yamaguchi E., Matsunaga K., Aramaki H., Horiuchi T.;				
RA	"Cloning and nucleotide sequences of NADH-pyridine oxidoreductase				
RT	P-450cam and putidaredoxin gene (cam) involved in cytochrome				
RT	P-450cam hydroxylase of Pseudomonas putida.";				
RL	J. Biochem. 106:831-836(1989).				
RN	[3]				
RP	SEQUENCE.				
RX	MEDLINE: 83030788.				
RA	Hanlu M., Armes L.G., Yasunobu K.T., Shastri B.A., Gunsalus I.C.;				
RT	"Amino acid sequence of the Pseudomonas putida cytochrome P-450. II.				
RT	Cytogen bromide peptides, acid cleavage peptides, and the complete				
RT	sequence.";				
RL	J. Biol. Chem. 257:12664-12671(1982).				
RN	[4]				
RP	X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).				
RX	MEDLINE: 86143817.				
RA	Finzel B.C., Weber P.C., Hardman K.D., Saleme F.R.;				
RT	"Structure of ferricytochrome c' from Rhodospirillum rubrum at				
RT	1.67-A resolution.";				
RL	J. Mol. Biol. 186:627-643(1985).				
RN	[5]				
RP	X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).				
RX	MEDLINE: 86059514.				
RA	Poulos T.L., Finzel B.C., Gunsalus I.C., Wagner G.C., Kraut J.;				
RT	"The 2.6-A crystal structure of Pseudomonas putida cytochrome P-450.";				
RL	J. Biol. Chem. 260:16122-16130(1985).				
RN	[6]				
RP	X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).				
RX	MEDLINE: 98019009.				
RA	Schlichting I., Jung C., Schulze H.;				
RT	"Crystal structure of cytochrome P-450cam complexed with the (1S)-				
RT	camphor enantiomer.";				
RL	FEBS Lett. 415:253-257(1997).				

FN	[17]		
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).		
RX	MEDLINE; 98313255.		
RA	VIKAKOVIC M., SILGAR S.G., LI H., POULOS T.L.;		
RT	"Understanding the role of the essential Asp251 in cytochrome p450cam		
RT	using site-directed mutagenesis, crystallography, and kinetic solvent		
RT	isotope effect.";		
RL	Biochemistry 37:9211-9219(1998).		
RN	[8]		
RP	STRUCTURE BY NMR.		
RX	MEDLINE; 97459726.		
RA	Mouro C., Bordon A., Simeoneaux G., Jung C.;		
RT	"1H-NMR study of diamagnetic cytochrome P450cam: assignment of heme		
RT	resonances and substrate dependence of one cysteinate beta proton.";		
RT	FEBS Lett. 414:203-208(1997).		
CC	-1- FUNCTION: INVOLVED IN A CAMPHOR OXIDATION SYSTEM.		
CC	-1- CATALYTIC ACTIVITY: CAMPHOR + PUTIDAREDOXIN + O(2) = 5-EXO-		
CC	HYDROXYCAMPHOR + OXIDIZED PUTIDAREDOXIN + H(2)O.		
CC	-1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as their content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
DR	EMBL; M12546; AAA25760.1; -;		
DR	EMBL; D00528; BAA00412.1; -;		
DR	PIR; A25660; OAPSCP.		
DR	PDB; 2CPP: 15-APR-91.		
DR	PDB; 3CPP: 15-APR-91.		
DR	PDB; 4CPP: 15-JUL-91.		
DR	PDB; 5CPP: 15-JUL-91.		
DR	PDB; 6CPP: 15-JUL-91.		
DR	PDB; 7CPP: 15-JUL-91.		
DR	PDB; 8CPP: 15-JUL-91.		
DR	PDB; 1CP4: 15-JAN-93.		
DR	PDB; 2CP4: 15-JAN-93.		
DR	PDB; 3CP4: 15-JAN-93.		
DR	PDB; 4CP4: 15-JAN-93.		
DR	PDB; 5CP4: 16-SEP-98.		
DR	PDB; 6CP4: 16-SEP-98.		
DR	PDB; 1NOO: 08-MAR-96.		
DR	PDB; 1PHA: 31-OCT-93.		
DR	PDB; 1PHB: 31-OCT-93.		
DR	PDB; 1PHD: 31-OCT-93.		
DR	PDB; 1PHE: 31-OCT-93.		
DR	PDB; 1PHF: 31-OCT-93.		
DR	PDB; 1PHG: 31-OCT-93.		
DR	PDB; 1AKD: 19-NOV-97.		
DR	PFAM; PF00067; P450_1.		
DR	PROSITE; PS00086; CYTOCHROME_P450_1.		
KW	Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;		
KW	3D-structure.		
FT	INIT MET	0	
FT	BINDING	357	HEM.
FT		55	MISSING (IN REF. 3).
FT	CONFLICT	276	E -> Q (IN REF. 3).
FT	CONFLICT	361	H -> S (IN REF. 3).
FT	CONFLICT	407	D -> N (IN REF. 3).
FT	TURN	16	
FT	TURN	20	
FT	HELIX	22	
FT	STRAND	23	
FT	TURN	28	
FT	TURN	32	
FT	HELIX	34	
FT	HELIX	38	
FT	TURN	48	
FT	STRAND	53	
FT	HELIX	58	

FT	STRAND	62	65
FT	HELIX	68	76
FT	TURN	78	80
FT	STRAND	81	82
FT	TURN	83	84
FT	HELIX	90	95
FT	TURN	99	102
FT	TURN	105	107
FT	HELIX	108	119
FT	HELIX	121	145
FT	TURN	146	146
FT	STRAND	147	149
FT	HELIX	150	153
FT	TURN	154	156
FT	HELIX	157	167
FT	TURN	168	168
FT	HELIX	171	185
FT	HELIX	193	213
FT	HELIX	219	224
FT	TURN	225	225
FT	STRAND	227	228
FT	TURN	229	230
FT	STRAND	231	232
FT	HELIX	235	250
FT	TURN	251	252
FT	HELIX	253	266
FT	HELIX	268	276
FT	HELIX	278	291
FT	TURN	292	292
FT	STRAND	295	295
FT	STRAND	297	301
FT	STRAND	305	307
FT	TURN	308	309
FT	STRAND	310	312
FT	TURN	314	315
FT	STRAND	317	320
FT	HELIX	322	324
FT	TURN	325	327
FT	TURN	329	331
FT	TURN	335	336
FT	TURN	340	341
FT	TURN	349	350
FT	HELIX	353	355
FT	TURN	358	359
FT	HELIX	360	377
FT	STRAND	382	383
FT	TURN	385	386
FT	STRAND	391	392
FT	STRAND	396	396
FT	STRAND	398	399
FT	STRAND	403	405
FT	HELIX	408	410
EQ	SEQUENCE	414 AA;	46538 MW; ECA610293A906207 CRC64;

Query Match	99.8%	Score 2176:	DB 1:	Length 414:
Best Local Similarity	99.8%:	Pred. No. 9e-159:		
Matches 413:	Conservative 1:	Mismatches 0:	Indels 0:	Gaps 0:
QY 1	TTETIQSANLAPLPVPHVPEHLVDFEDMYNPSNLSAGVOEAMVLQESNVDPLVYTRCNG	60		
Db 1	TTETIQSANLAPLPVPHVPEHLVDFEDMYNPSNLSAGVOEAMVLQESNVDPLVYTRCNG	60		
QY 61	GHWIATRCQLIREAVEDYRHFSSCECFIPREAGEAYDITPSMDPPEQROFALANQYVG	120		
Db 61	GHWIATRCQLIREAVEDYRHFSSCECFIPREAGEAYDITPSMDPPEQROFALANQYVG	120		
QY 121	MPVYDKLENRIQELACSLIESLRPOGQCNPEEDVAEPPRIITFMLLAGLPEDDIPHLKYL	180		
Db 121	MPVYDKLENRIQELACSLIESLRPOGQCNPEEDVAEPPRIITFMLLAGLPEDDIPHLKYL	180		
QY 181	TDQMTRPDGSMTFEAKKALDYLLPIIEQRROKPGTDAISIVANGQNGRPITISDEAKR	240		



Db	181	TDQMRPDSMTPEAKALDYDLPIIEQRQRKPGTDAISVANGVGNRPITSDEAKR	240
QY	241	MGGLLVGGDLTVVNFPSFSMEFLAKSPBHRQELIERPERIPAAACELLRRFSVLADGRI	3000
Db	241	MGGLLVGGDLTVVNFPSFSMEFLAKSPBHRQELIERPERIPAAACELLRRFSVLADGRI	3000
QY	301	LTSDFEFGVGLKKGDQILLPQMLSGIDERNKCPMIVDSRQKVSHTTGHSHCLGQ	3600
Db	301	LTSDFEFGVGLKKGDQILLPQMLSGIDERNKCPMIVDSRQKVSHTTGHSHCLGQ	3600
QY	361	HLARREIIVTLKEMUTRIIPDSIAPGAOIOHKSGISVGVALLPMPMPATTKAV	414
Db	361	HLARREIIVTLKEMUTRIIPDSIAPGAOIOHKSGISVGVALLPMPMPATTKAV	414

ID	YVAG_RH1SN	STANDARD:	PRT:	414 AA.
AC	Q53215:			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	15-DEC-1998 (Rel. 37, Last annotation update)			
DE	PROBABLE CYTOCHROME P450 YVAG (EC 1.14.14.-).			
GN	YVAG.			
OS	Rhizobium sp. (strain NGR234).			
OG	Plasmid sym PNGR234a.			
OC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;			
OC	Rhizobiaceae; Rhizobium.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE: 97305956.			
RA	Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,			
RA	Perret X.;			
RT	"Molecular basis of symbiosis between Rhizobium and legumes.";			
RL	Nature 387:394-401(1997).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE: 96389014.			
RA	Freiberg C., Perret X., Broughton W.J., Rosenthal A.;			
RT	"Sequencing the 500-kb GC-rich symbiotic replicon of Rhizobium sp.			
RT	NGR234 using dye terminators and a thermostable 'sequenase': a			
RT	beginning.";			
RL	Genome Res. 6:590-600(1996).			
CC	-I- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE			
CC	MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED			
CC	COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.			
CC	-I SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation at			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and that this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL: Z68303: CAA92422.1: -			
DR	EMBL: AE000101: AAB91895.1: -			
DR	HSSP: P00183: 3CP4.			
DR	PFAM: PF00067: P450_2.			
DR	PROSITE: PS00086: CYTOCHROME_P450; FALSE NEG.			
KW	Hypothetical protein; Oxidoreductase; Monooxygenase;			
KW	Electron transport; Heme; Plasmid.			
FT	BINDING 364 HEME (BY SIMILARITY).			
SO	SEQUENCE 414 AA; 45810 MW; 888598E99315BB84 CRC64;			

	Query Match	27.1%	Score 591.5;	DB.1:	Length 414;
	Best Local Similarity	34.1%;	Pred. No. 5.9e-38;		
	Matches 142; Conservative	65;	Mismatches 171;	Indels	39; Gaps
QY	13 PLPRVPEHLYEDFMNPSNLIA---CYQEAMVALQESNPVDLW----TRCNGGHMIA	65			
	:               :	:	:	:	:         :

```

Db      13  PIPDHVPALVRHRSLSIFSPGMATPTPNDDPAAVAACVHDSDPPIFFYSPSNTROSGRWVI  72
Oy      66  TRGQLIREAYEDYRHFSSECPFIPREAGEAYDFTPTSMDDPEOROFALANQVGMFVVD  125
Db      73  TRARQORVLEDTETFSHSRISFASALCEHMPVPIPLEDPAHGVAFLANLPFSRRVL  132
Oy      126  KLENIIOELACSLIESL-RRGQGNFTEDYAEPPPIRIFMLLAGLPE-----EDI  174
Db      133  ALEPTIHARACALIDCLAKETSCDYMKDFALPPTFSVFLSFLGLSORREVLVGVSDL  192
Oy      175  PLKLTLDQMRPGSGMFAFAKALDYLLPIIEQ-----RROKPGDAISIVANGGVNG  230
Db      193  LH-----GN--AKRRRAAARSVAVFIDEMAANRRKSPAYDFMTYVQAKIEG  237
Oy      231  RPTSDAEKRMCGLLVGLDVTYVNFLSFSMEFLAKSPENROELIERPERIAPACEILR  290
Db      238  RSLTFEEBVRIGVLFVLVAGDLTVAAALIGDFMAYLARPNKHQELLRNPARGLAAEBELR  297
Oy      291  RPSLVADORILTSDEYFNGVOLKKGDQILPQMLSGLDERKNACPMHVDTSROKVSHTTF  350
Db      298  AYSTVOIIRVATKDIIEFGCVPIRGCDVYSCAPMIANRPPSEFKCPNTIDLARONQHTAF  357
Oy      351  GHGSHLGCQHLARREITVTLKEMNLTTRIPDSIAPG-AQIOHKSGIAYGOVALPLYW  406
Db      358  GYGPHTLCGAILHAREIYIGLRREMLARIPARIRIEGAPLTH-GGAYFGGSIKITLW  413

```

RESULT	3
YH85_MYCMU	
ID	YH85_MYCTU STANDARD: PRT: 393 AA.
AC	053936;
DT	15-FEB-2000 (Rel. 39, Created)
DT	15-FEB-2000 (Rel. 39, Last sequence update)
DT	15-FEB-2000 (Rel. 39, Last annotation update)
DE	PUTATIVE CYTOCHROME P450 RV1785C (EC 1.14.--).
GN	RV1785C OR MTY049.07C.
OS	Mycobacterium tuberculosis.
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=H37RV;
RX	MEDLINE: 98295987.
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA	Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA	Holmesby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA	Oliver S.S., Osborne J., Quail M.A., Rajadream M.A., Rogers J.,
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA	Taylor K., Whitehead S., Barrall B.G.;
RT	"Deciphering the biology of Mycobacterium tuberculosis from the
RL	complete genome sequence."
RL	Nature 393:537-544(1998).
CC	-1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation --
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	
DR	EMBL: AL022021; CAU17707.1; --
DR	TUBERCULIST; RV1785C; --
DR	PFAM: PF00067; P450; 2.
DR	PROSITE: PS00086; CYTOCHROME_P450; 1.
KW	Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.
FT	BINDING 342 342 HEME (BY SIMILARITY).
SO	SEQUENCE 393 AA: 43541 MW: 86CDID508471FAB8 CXC64;

Query Match 17.3%; Score 377; DB 1; Length 393;  
 Best Local Similarity 28.7%; Pred. No. 1.3e-21;  
 Matches 110; Conservative 70; Mismatches 183; Indels 20; Gaps 10;

CC 31 PSLNLSAGVQEMANVLOSNNVDLYWTRCNGHMTATRGQLIREVEDYRHHSSCEPFIIPR 90  
 DB 23 PMAYDRGV--GMYTLRAG--PVVF--MNGMYVL--TRREDLALRNPKYFSSSKALQP- 74  
 QY 91 EAGEAVYFIPTSMDEPQROFRALANQVGMVVDKLENRIQELACSLIESLRPGOCNF 150  
 DB 75 -PGRLPLVPLAFDPPEHTRRRLQRYFSPALSKALPSLRRTVAMIDIAIRGCEA 133  
 QY 151 TEDYAEPPRIEMLAGLPREDIPHKYLTQDQ---TRPDGSMTEAKNEALYDYLIP 206  
 DB 134 MADLANFEPQLFLYGLPLEDRDLIGMWDVIAIAMSRRPHPEADVAARLELEYTA 193  
 QY 207 IIEORRQKPGDAISYANGVNGRPITSDAKRMCGLLVGLDYVNFLESMELFLAK 266  
 DB 194 MVAERRRNPGDVLISQV---QIGEDPLSEIEVLGLSHLLIAGLDVYAAVGFSLLELAR 250  
 QY 267 SPEHROELIERPERIPACCELLR-RFSLVADGRILTSDEYFHVOLKKGOILLPQMLS 325  
 DB 251 RPOLRAALRNPKQIRFIEIYVLESAPVAPVTEPTVYGMTLPAGSPVLCMAAV 310  
 QY 326 GLDERKNACPMHVDPSROKVSHTTGHGSHLCLGQHLARREIIVTLKEMWLRIPDFSIA 385  
 DB 311 NRQDSAMSTDELVMGDKVHRHMGFGGPHRCGSHLARLELTLLVGBMLNQIDFELAP 370  
 QY 386 --GAQIOHKSGIVSGVALPLW 406  
 DB 371 DYAEIRFPSPKSA-LKNLPLRW 392

RESULT 4  
 CPXM\_BACME STANDARD; PRT; 410 AA.

AC Q06069;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-FEB-2000 (Rel. 39, Last annotation update)  
 DE CYTOCHROME P450(MEG) (EC 1.14.99.-) (STEROID 15-BETA-HYDROXYLASE)  
 DE (STEROID 15-BETA-MONOXYGENASE).  
 GN CYP106A2.  
 OS Bacillus megaterium.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 13368;  
 RX MEDLINE; 94049677.  
 RA Rauschenbach R., Isernhaugen M., Noeske-Jungblut C., Boldol W.,  
 RA Stewart G.;  
 RT "Cloning sequencing and expression of the gene for cytochrome  
 RT P450meg, the steroid-15 beta-monoxygenase from Bacillus megaterium  
 RT ATCC 13368.";  
 RL Mol. Genet. 241:170-176(1993).  
 RN [2]  
 RP CHARACTERIZATION.  
 RC STRAIN-ATCC 13368;  
 RX MEDLINE; 79194051.  
 RA Berg A., Ingelman-Sundberg M., Gustafsson M.;  
 RT "Purification and characterization of cytochrome P-450meg.";  
 RL J. Biol. Chem. 254:5264-5271(1979).  
 RN [3]  
 RP CHARACTERIZATION.  
 RC STRAIN-ATCC 13368;  
 RX MEDLINE; 82091079.  
 RA Berg A., Rafter J.J.;  
 RT "Studies on the substrate specificity and inducibility of cytochrome  
 RT P-450meg.";  
 RL Biochem. J. 196:781-786(1981).  
 CC -1- FUNCTION: HAS THE CAPACITY TO HYDROXYLATE CERTAIN STEROIDS IN THE  
 CC 15-BETA POSITION.

CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; Z21972; CA97985.1; -.  
 DR PIR; S32216; S32216.  
 DR PIR; S39924; S39924.  
 DR PFAM; PF00067; P450.1.  
 DR PRINTS; PR00359; BP450.  
 DR PROSITE; P500066; CYTOCHROME\_P450; 1.  
 KM OXIDOREDUCTASE; MONOOXYGENASE; ELECTRON TRANSPORT; MEMBRANE; HEME.  
 FT BINDING 355 355 HEME (BY SIMILARITY).  
 SQ SEQUENCE 410 AA; 46955 MW; 9FDD9CF2E0F810B CRC64;

Query Match 17.1%; Score 372; DB 1; Length 410;  
 Best Local Similarity 30.1%; Pred. No. 3.2e-21;  
 Matches 109; Conservative 53; Mismatches 180; Indels 20; Gaps 7;

CC 63 WIATRGQILREAYEDYRHHFSE-----CPPIREAGEAYDFIP-TSMDEPQROFRALA 115  
 DB 44 WNVFYEDYKRVLSYDYKHFSSVRRKRTISVGTDBEESVPEKIDITSSDPDRHKKRSL 103  
 QY 116 NVVGMVVDKLENRIQELACSLIESLRPGOCNFTEDVAPPPRIEMLAGLPREDIP 175  
 DB 104 AAFTPRSLQNWEPRIQELADELIGMDGTEIDIVASLSPRIIYVADLMGVPSKRL 163  
 QY 176 HLKLTIDQMTDPGSMFAP-----AKELVYLLPIIIEORRQKGTDAISVANGV 228  
 DB 164 LFKRWVDTLFLPDREREKOEVDKLOYAKERYOYLPIYVQKRLNPADIIISDLKSEV 223  
 QY 229 NGRPITSDAKRMCGLLVGLDYVNFVNFSPM-EFLAKSPENROELIERPERIPACCE 287  
 DB 224 DGEFTDDEVYRTMLLIGACVETTSHLANSFSLYDKDEVQELHENDLVPAVEE 283  
 QY 288 LLR-RFSLVADGRILTSDEYFHVOLKKGOILLPQMLSGDERKNACPMHVDPSR-QKV 345  
 DB 284 MLRFRFMLIKIDRTYKEDNDLGLVELKEGDSVVVWMSAANDDEMFDPFLINIRPNK 343  
 QY 346 SHTFEGHSHCLGQHLARREIIVTLKEMWLRIPDFSIAQAQQRK---SGIVSGVAL 402  
 DB 344 KHLTFEGNGPHFCIGAPLARLEAKTALTAFLKFKHIAVPSFOLENLTDATGQTLTSL 403  
 QY 403 PL 404  
 DB 404 PL 405

RESULT 5  
 YJTB\_BACSU STANDARD; PRT; 396 AA.

AC Q34374;  
 DT 15-FEB-2000 (Rel. 39, Created)  
 DT 15-FEB-2000 (Rel. 39, Last sequence update)  
 DT 15-FEB-2000 (Rel. 39, Last annotation update)  
 DE PUTATIVE CYTOCHROME P450 YJTB (EC 1.14.-.-).  
 GN YJTB.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-168;  
 RA Rivolta C., Soldo B., Lazarevic V., Joris B., Manuel C., Karamata D.;  
 RT "A 35.7 kb DNA fragment from Bacillus subtilis chromosome containing a  
 RT putative 12.3 kb operon involved in heuronic catabolism and a  
 RT perfect catabolite-responsive element.";

R.L Submitted (JUL-1997) to the EMBL/Genebank/DDBJ databases.

CC -I SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement ([see http://www.isb.ch/announce/](http://www.isb.ch/announce/))  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

CC -----

DR EMBL; AF015825; AAC46317.1; -.  
DR EMBL; Z29110; CAB13078.1; -.  
DR SUBTLIST: BG131395; XJTB.  
DR PRAM; PF00067; p450; 1.  
DR PRINTS; PRO0359; BP450.  
DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme-  
FT BINDING 349 349 heme (BY SIMILARITY).  
SQ SEQUENCE 396 AA; 44990 MW; 9A89C12613DBCFB CAC64;

Query Match 16.8%; Score 366; DB 1; Length 396;  
Best Local Similarity 30.4%; Pred. No. 8.8e-21;  
Matches 100; Conservative 54; Mismatches 157; Indels 18; Gaps 7;

OY 71 IREAVEDRYHSSSECPPIPREAGEAYDTIPMSDPDEQROFALANOVGMVVDKLENR 130  
Db ::::| | | : : : : : ||||| : : : : : : : :  
Oy 56 VKVVGDKELEFS--C--MPQQSTSGNST-IIMDDPKHTKIKSYNNKAFTPVMKQMEPR 111  
Db 111| | | : : : : : : : : : : : : : : : : : :  
Oy 131 IOELACLILESLRRGQCNGTFDYAEPPRIIRFMLLAGLPREDIHLYLTQM----- 185  
Db 112 IOETDELLOKRQGSRSEDVLHDVSFYRPVPVIYSIELGVPSAHMDOFAWSDLVSTPDK 171  
Oy 186 --RPDGSMTFAE---AKFEALDYLIPIIEORAKOGCTGAISVANQVVGRTIDEAKR 240  
Db 172 KSEEAFKAFLEBRDCCEBLAFAPFGIGIIKKRKNEODITSLVAEELEGKLSBELLR 231  
Oy 241 MCGILLVGGLDVNVNLFSTSMEFLAKSPENHOELIERERIPACEBELLR-RFSLVADGR 299  
Db 232 FCTLTVLGANGNETTTMLNISAWYSILETPGYVELRSHPQLMQAVEBALRFAPAVALRL 291  
Oy 300 ILTSYEFRHGVOVKRGDQDLLPLQMLSGDERKNNAACPNIHVDFSKRVSHTFPHGSHCLTG 359  
Db 292 IAKRTELETGHLIKKGDMVLAFVASANDREAEDFRPMHIRHNPNHAIFGHIHFCLG 351  
Oy 360 OHLARREIVTLKEMILTTRIPDE---SIAP 385  
Db 352 APLARLEANIALTSLISAFFHMESYSTIP 380

RESULT 6  
CPXM\_BACSU  
ID CPXM\_BACSU STANDARD: PROT: 405 AA.  
DI P27632:  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 15-FEB-2000 (Rel. 33, Last annotation update)  
DE CYTOCHROME P450 109 (EC 1.14.-.-) (ONF405).  
GN CYP109  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillius/Clostridium group;  
NC Bacillius/Staphylococcus group; Bacillius.  
RN [1]  
RC SEQUENCE FROM N.A.  
RX STRAIN-WZ3:  
RA MEDLINE: 91192601.  
RA Ahn K.S., Wake R.G.;  
RT "Variations and coding features of the sequence spanning the  
RT replication terms of Bacillus subtilis 168 and WZ3 chromosomes.",  
RL Gene 98:107-112(1991).  
CC -I FUNCTION: CYTOCHROMES P450 ARE A GROUP OFHEME-THIO-LATE  
CC MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED

```

CC COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcements).
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M24523; AAA22720.1; -.
CC DR HSSP; 000441; 10XA.
CC DR PRFM; PF00067; P450; 1.
CC DR PRINTS; PR00359; Bp450.
CC DR PROSITE; PS00086; CYTOCHROME_P450; 1.
CC DR Oxidoreductase; Monooxygenase; Electron transport; Heme.
CC FT BINDING 351 351 HEME (BY SIMILARITY).
CC SO SEQUENCE 405 AA; 45845 MW; 1331D5BEA74E3C04 CRC64;
CC -----
Query Match 16.6%; Score 361.5; DB 1; Length 405;
Best Local Similarity 25.2%; Pred. No. 2e-20;
Matches 113; Conservative 72; Mismatches 169; Indels 95; Gaps 10;
QY 1 TTTTQSN-----ANLADLP- HYPEHLVFEDMTNPSNLSAGV-----QEAVALQESN 49
Dd 2 TNOTARRSKKKERYANLIMEELSEKDLFPPIYDKLRRESVRYDPLRDCMDYFK--- 58
QY 50 VPDLVWTCNGNGHWTATGOLIREAYEDYRHHSSCEPPIPRAGENAYDFIPTSMPPEDR 109
Dd 59 -----YDDVQVFLNPKLFSSKRGIGQTESILT-MDPKHT 92
QY 110 QFRALANQVGMPPVYDKLENRIOLACSLIESLRPOGONCFEDYAEPPRIRIFMLAGL 169
Dd 93 KLRALVSNAFPKPAVKOLETRIKDVTATLQEAROKSTIIDIEDAGPLVYIIMELGA 152
QY 170 PEEDIPHLKYLTDMQTRPDSWTFEAKEA-----LYDYLIPIIDR 211
Dd 153 PIEDHLLIKTYSDVLY-----AGAKDSSDKAAVDWVHNRDGHAFLDYDFRDLISK 204
QY 212 RQKPTDIAISIVANQVNGRPTSDPAKRMGLLVGLGDIYVYNLSPSMELASPEIR 271
Dd 205 RAEPEDLMTMLQAEIDGEYLTGEOLIGFCLLVAGNETTTLIANAVRLLTDSVQ 264
QY 272 QELLERPERIPAAECBELLRFSLY-ADGRILTSDYEFHGVOLTKGQDILLPOMLSGLDR 330
Dd 265 QOVRNTDNVANVIEETLRYISPVQAIGRVATEDIELGCVFKKSSVYSWIASANRDE 324
QY 331 KNAECMAHVDFSROKVSHTTFGHSHLCIGOLHARREIIVTK-----EWL 375
Dd 325 KFKCPDCKEIRDPSPYRPHLSFGFICGAPLARLEANIASSLSMSACIEKAHDEKL 384
QY 376 TRIPDFSIAPGAQIOHKSGIVSGVALP 404
Dd 385 EAIPI-----SPFVFGVGRLLPV 400
RESULT 7
ID Y218_MYCTU STANDARD: PRT: 398 AA.
AC 053563;
DT 15-FEB-2000 (Rel. 39, Created)
DT 15-FEB-2000 (Rel. 39, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE PUTATIVE CYTOCHROME P450 RV3518C (EC 1.14.-.-).
GN RV3518C OR MTW023.25C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RC STRAIN=H37RV.

```

RX MEDLINE: 98295987.  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Baddock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter K., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,  
RA Taylor K., Whitehead S., Barrall B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence.";  
RL Nature 393:537-544(1998).  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AL022022; CA17755.1; -  
DR TUBERCULIST; RV3518C; -  
DR PRAM; PF00067; P450; 1.  
DR PRINTS; PR00359; BP450.  
DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
KM Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.  
FT BINDING 340 340 HEME (BY SIMILARITY).  
SQ SEQUENCE 398 AA; 44398 MW; BCF3C23ECB5767F CRC64;

Query Match 16.5%; Score 360; DB 1; Length 398;  
Best Local Similarity 28.2%; Pred. No. 2.5e-20;  
Matches 107; Conservative 69; Mismatches 169; Indels 34; Gaps 8;

OY 55 WTRCN-----GGHWIATRGQIREAYEDYRHFS-----ECPTIPREAGEAYDPIPT 101  
DB 26 WMRAQNVFVRNRLAASSTYQAVIDAERQPELFSNAGIRPDQPALP-----MMI 76  
OY 102 SMDPEQROPRALANOVGVNVDKLENRTQELACSLIESLRPGOCNFTEDYAEPPPIR 161  
DB 77 DMDPAHLRRLKLVNAGFTTRKRVKDKASIALCDTLIDACGECDFVNDLAAPMA 136  
OY 162 IFMLAGL-PEEDIPHLK-----YLTDMTRPDGSMTEFAEAKALYDYLPITTEORQ 213  
DB 137 VIGMGLVREQORMLFLRMDDLVTLFSHVSOEDPOITM-DATAAINDPFRATIAARRA 195  
OY 214 KPGTDAISIVANGOVNCRPITSDEAKRMGILLVGLDVTYVNFSLFSMEPLAKSPHROE 273  
DB 196 DPTDDLVSVLVSSVEVDEGRISDELVMEETLILIGDETTTRHFLSGTGEOLIRNRQMDL 255  
OY 274 LIERPERIPACEELRLRFSIVAD-GRILTSDYERHGVOLKKGQIILLPQMLSLDERKN 332  
DB 256 LQDRPSLLPGALIEMLMTAPVKNMCVRLTADTEFHGTALCAGKMLLESAFNEDAVP 315  
OY 333 ACPMHVDFSRQKSHTTFGHSHLCLGOHLARREIIVTLKEMLRIPDFSIAPGAQI--Q 390  
DB 316 CEPEKRFVQGNPNNSHLAFGFTGHCLGNOLARLELSLMTREVLRLRDLRLVADSVLPL 375  
OY 391 HKSGIVSGVQALPLWMDPA 409  
DB 376 RPAFNVSGLESMVPVFPFS 394

RESULT 8  
CPXF\_STRGO STANDARD; PRT; 402 AA.  
ID CPXF\_STRGO STANDARD; PRT; 402 AA.  
AC P18327;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE CYTOCHROME P450-SU2 (EC 1.14.-.-) (P450-CVB1) (CYP105B1).

GN CYP105B1 OR SUBC.  
OS Streptomyces griseolus.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycinaceae; Streptomycetaceae; Streptomyces.  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-32.  
RC STRAIN-ATCC 11796;  
RX MEDLINE: 90264332.  
RA Omer C.A., Lenstra R., Little P.J., Dean C., Tepperman J.M.,  
RA Lato K.J., Romesser J.A., O'Keefe D.P.;  
RT "Genes for two herbicide-inducible cytochromes P-450 from  
RT Streptomyces griseolus.";  
RL J. Bacteriol. 172:3335-3345(1990).  
CC -1- FUNCTION: METABOLISM OF A NUMBER OF SULFONYLUREA HERBICIDES.  
CC -1- INDUCTION: BY HERBICIDE.  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M32239; AAA26825.1; -  
DR PIR; B35401; B35401.  
DR HSSP; P23295; 2ROM.  
DR PRAM; PF00067; P450; 1.  
DR PRINTS; PR00359; BP450.  
DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
KM Oxidoreductase; Monooxygenase; Electron transport; Heme.  
FT INT MET 0 0  
FT BINDING 351 351 HEME (BY SIMILARITY).  
SQ SEQUENCE 402 AA; 44278 MW; E3B67F672C2609D CRC64;

Query Match 15.6%; Score 340; DB 1; Length 402;  
Best Local Similarity 28.9%; Pred. No. 8.6e-19;  
Matches 105; Conservative 63; Mismatches 165; Indels 30; Gaps 13;

OY 63 WINTRGQIREAYEDYRHFSSEC-----PTIPREAGEAYDPIPT--SMDPEQROPRALA 115  
DB 51 WLVTIRHODVRAVLGDPR-FSADAHRGTGFPFLTAGGREIIGTNPFLRMDDEHARLRRL 109  
OY 116 NOVGVNVDKLENRTQELACSLIESLRP-QGOCNFTEDYAEPPPIRIFMLAGLPREDI 174  
DB 110 TADFTYAKVEAMREPVQRLADDLVDRMTGRTSDADVTETALPLPSLVLTLLGVPYEDH 169  
OY 175 PHL-----KYLTDMTRPDGSMTEFAEAKALYDYLPITTEORQKPGTDAIS-IVANGOVN 229  
DB 170 AFQGERSRVLLTLRSPE---EVRAAODELLFYLARLARKRKRPDAIISRVLANGELD 226  
OY 230 GRPITSDEAKRMGILLVGLDVTYVNFSLFSMEPLAKSPHROELIERPRIPACEELL 289  
DB 227 DTQIAT-----MGRLLVAAHETTANMTALSTVLLENPQRLRLRAEPALVAGAVEEL 281  
OY 290 RRSIVADG--RLTTSYERHGVOLKKGQIILLPQMLSGDERKNCR--MHVDFSRQKY 345  
DB 282 RYLTIVANGVPRIATEVTLIGGRTIAGBEVLG--MISANRDAEVPFGDDDLVADAR 339  
OY 346 SHTTFGHSHLCLGOHLARREIIVTLKEMLRIPDFSIAPGAQI--Q 403  
DB 340 RHVAFGCVHQCIGQPLARVELDQIALETLLRLRDLRLAVPHEIPRGMALYGVHSLP 399  
OY 404 LVW 406  
DB 400 IAW 402

RESULT 9  
YM66\_MYCTU STANDARD; PRT; 428 AA.  
ID YM66\_MYCTU STANDARD; PRT; 428 AA.

AC 050696;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-FEB-2000 (Rel. 39, Last annotation update)  
DE PUTATIVE CYTOCHROME P450 RV2266 (EC 1.14.-.-).  
GN RV2266 OR MTCY339.44C.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H37RV.  
RX MEDLINE; 98295987.  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,  
Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,  
Hornsby T., Jagels K., Kitchin A., McLean J., Moule S., Murphy L.,  
Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
Rutler S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,  
Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence."  
RL Nature 393:537-544(1998).  
CC -I- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Z71163; CAB00969.1; -.  
DR HSSP; P33006; 1CPT.  
DR TUBERCULIST; RV2266; -.  
DR PFAM; PF00067; P450.1.  
DR PRINTS; PR00359; BP450.  
DR PROSITE; PS00086; CYTOCHROME\_P450; FALSE\_NEG.  
KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.  
FT BINDING 379 379 HEME (BY SIMILARITY).  
SQ SEQUENCE 428 AA; 47824 MW; 76B1F3C5AE348591 CRC64;

Query Match 15.5%; Score 338.5; DB 1; Length 428;  
Best Local Similarity 26.2%; Pred. No. 1.2e-18;  
Matches 117; Conservative 72; Mismatches 188; Indels 69; Gaps 17;

QY 5 IOSNANLAPRPHVPEHLYVDFDMYNSNLSAGVOEAMAVIQE-----SNVPLV 54  
DB 7 IATRYNGTP-PPEVP--IADIEL-----GSLDFWALDDVDGAFATLRREAPISF 54  
QY 55 WTR-----CNGGHMIATRCQLREAYEDY---RH---FSS-----ECPTIPREA 92  
DB 55 WPTIELPGFVANGNGMALTK-----YDVFYARHRHDFISSYNTITINQTPLEAYF 107  
QY 93 GEAVDFITSDPPEQROFRALANOVGMVYVDKLENRIQELACSLIESL---RPQGCN 149  
DB 108 GSM-----IVLDDPRHQRISRIVSRAFTPKVARIETEAIVRDAHRLVSSMLANNDDROAD 162  
QY 150 FTEDVAFEPFIRIFMLLAGLPEEDIPHLAKYLDDMT---RPGDSMTFAEAKALYD---Y 203  
DB 163 LVSELAGLPLQDITCDMMGIRPADHQRIHMTNVLGFGDPLAVDFEFQVNSADIGAY 222  
QY 204 LPIIEQRORPGTDAISIVANGVNGRPITSDEAKRMCGLLVGLTGVVFLSFSMEF 263  
DB 223 ATALAEADRVRNHHDJLSSIVLEAEVDGRSLRSREIASFILLVAGNETTRNAITHGVLA 282  
QY 264 LAKSEHROELIERPERI-PAACELLRRFS-LVADGRIILTSYDFHGVOLKKGQDQILP 321  
DB 283 LSRVPEQRDRMWSDFGLAPFAVEIIVMASPVVYMRRTLIQDILRGTAKMAAGKVSIM 342

QY 322 OMLGLDERKNKPMHVDYFSROKVSHTTF-GHSHLACLOHARREITVTLKEMLTRIPD 380  
DB 343 YCSANRDSKRADPTPTFLANPNPHLFGGGGAFICLGANLAREINVARDELAROMPD 402  
QY 381 FSIAPGAQIQHKSIGVSGVALPLVW 406  
DB 403 V-VAITEEPARLSQFIHGIKTLPTVW 427

RESULT 10  
CPXL\_BACME  
ID CPXL\_BACME STANDARD; PRT; 410 AA.  
AC P14762;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE CYTOCHROME P450(BM-1) (EC 1.14.14.1).  
GN CYP106.  
OS Bacillus megaterium.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-25.  
RC STRAIN-ATCC 14581;  
RX MEDLINE; 90089408.  
RA He J.S., Ruettinger R.T., Liu H.-M., Fulco A.J.;  
RT "Molecular cloning, coding nucleotides and the deduced amino acid  
RT sequence of P-450BM-1 from Bacillus megaterium."  
RL Biochim. Biophys. Acta 1009:301-303(1989).  
RN [2]  
RP SEQUENCE OF 1-25 FROM N.A.  
RX MEDLINE; 95355495.  
RA He J.S., Liang Q., Fulco A.J.;  
RT "The molecular cloning and characterization of Bm1p1 and Bm1p2  
RT proteins, putative positive transcription factors involved in  
RT barbiturate-mediated induction of the genes encoding cytochrome  
RT P450BM-1 of Bacillus megaterium."  
RL J. Biol. Chem. 270:18615-18625(1995).  
CC -I- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE  
CC MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED  
CC COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.  
CC -I- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X16610; CA34612.1; -.  
DR EMBL; S79230; AAC60495.1; -.  
DR PIR; S07764; O4865M.  
DR HSSP; P33006; 1CPT.  
DR PFAM; PF00067; P450.1.  
DR PRINTS; PR00359; BP450.  
DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.  
FT BINDING 356 356 HEME.  
SQ SEQUENCE 410 AA; 47460 MW; C9AE293EF6745387 CRC64;

Query Match 15.4%; Score 335.5; DB 1; Length 410;  
Best Local Similarity 27.7%; Pred. No. 1.9e-18;  
Matches 96; Conservative 57; Mismatches 170; Indels 23; Gaps 7;

QY 63 WIATGQLIREYEDYRRFSSF---CPPIPRAGALYDIP---TSDPPEQROFRALA 115  
DB 45 WNVFYEYHVKOVLNVYDFSSDGGRTTIFGVNSKRKSTSTPTNTJLNDPPHRRARSIL 104  
QY 116 NOVGMVYVDKLENRIQELACSLIESLRPGCCNTEDEYAEFPFIRIFMLLAGLPEEDIP 175

```
Db 105 AAAFPSPSLKMEPRRIQIAADLVEAIOKNTINIVDLSSPPSLVIADLFGVPVKDRY 164
QY 176 HLKYLTPDMTPRDCSMTFAEAKE-----ALYDYLPIIEORROKPGTDAISIVANGQV 228
Db 165 QFKWVIVLFPYDQERNEIEQEKORAGAEYFQYLYPIVIEKSNLSDDLISDLIAEYV 224
QY 229 NGRPITSDEAKRMGCLLVGLDFTVNFVLF-SFSMEFLAKSPERHQLIERPERIPAAEC 287
Db 225 DGEFTDEEIVHATMLLIGAGVETTSIAINMFYSFLYDDKSLXSELRNNRELAPKAVEE 284
QY 288 LLR-RFSLVADGRILTSDEYEHGVOLKKGDOILLPQMLSGDERKNACPMHVDSCR-QKV 345
Db 285 MLKRFHHSRRDRVKONELLVGLKKGDVYIAMMSACNDEMFPESVDIHRPTNK 344
QY 346 SHTFGHSLCGLQHLARREIIVTLKEMLTR-----IPDFSIAIP 385
Db 345 KHLFGNGPHRCGLAPLARLEMKIILAEFLAEFHSIEPFELEP 390

RESULT 11
CPKX_SACER STANDARD; PRT; 405 AA.
ID CPKX_SACER STANDARD; PRT; 405 AA.
AC P33271;
DF 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CYTOCHROME P450 107B1 (EC 1.14.-.-) (P450C11B1).
GN CYP107B1.
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Pseudonocardiales; Pseudonocardiaceae;
OC Saccharopolyspora.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 172-203 AND 302-321.
RC STRAIN=NRRL 2338;
RX MEDLINE; 92121109.
RA Andersen J.F., Hutchinson C.R.;
RT "Characterization of Saccharopolyspora erythraea cytochrome P-450
RT genes and enzymes, including 6-deoxyerythronolide B hydroxylase."
RL J. Bacteriol. 174:725-735(1992).
CC -1- FUNCTION: NOT KNOWN, PROBABLY INVOLVED IN THE CATABOLISM OF
CC OCTANE AND GUAIACOL. IT DISPLAYS A WEAK ACTIVITY IN THE O-
CC DEALKYLATION OF 7-ETHOXYCODYMARIN.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M83110; AAA26483.1; -
CC DR PIR; B42606;
CC DR HSSP; 000441; 10XA.
CC DR PFAM; PF00067; P450.1.
CC DR PRINTS; PR00359; BP450.
CC DR PROSITE; PS00086; CYTOCHROME_P450.1.
CC DR Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.
CC FT BINDING 352 352 HEME (BY SIMILARITY).
CC SEQUENCE 405 AA; 45238 MW; 71C93CEC1FDC33FD CRC64;
SO
```

```
Query Match 15.3%; Score 333.5; DB 1; Length 405;
Best local similarity 25.6%; Pred. No. 2.7e-18;
Matches 107; Conservative 74; Mismatches 184; Indels 53; Gaps 10;
```

```
QY 18 VPEHLVFD-----FDMYNPSNLASGVDEAMAVLQESNPDLWTRCNG-GHMIATRGQILIR 72
Db 6 VPDLAEFDDAFADQRHH-----RYARMEPEYQRI--RTVNLDMALITRIYEDVK 53
QY 73 EAYEDYRHFSSECPFIPRAGEAYDIPTSM-----DPEQRQFR 112
```

```
Db 54 QALD-----PRIAKDGRTOOILIEKRLADAERRRGFSPDLGPMMLNTDPPDHRRLR 105
QY 113 ALANQVGMVPVXDLLENRIQELACSLIESLRPOGQCNFTEDYAPPRIRIFMLLAGPEE 172
Db 106 KLVVKAFTARRVEGLRPIRIQITDDLRLAGRSEVLDIEFAPRLPIYVISELGEVDS 165
QY 173 DIPHLKYLTPDMTPRDCSMTFAEAKE--ALYDYLPIIEORROKPGTDAISIVANGVNG 230
Db 166 RRDDFRSMTNVLV--DSSQPEAQASVAMVEYITELIAKRTPEPGDDLTALEAVEDG 223
QY 231 RPIISDEAKRMGCLLVGLDFTVNFVLF-SFSMEFLAKSPERHQLIERPERIPAAEC 287
Db 224 DRUSEGELIAMVFLLVAGHETVNLIGNCVLSLGNPDOLAALRNPSLPGAIETELR 283
QY 291 RFLSVADG--RLTSDEYEHGVOLKKGDOILLPQMLSGDERKNACPMHVDSCRQKSH 348
Db 284 YESPVANGTFRHTAEAVRFGDVVPESELVAVGAANRGEREDPDRTRETTGHW 343
QY 349 TFGHSHLCLGQHLARREIIVTLKEMLTRIPDFSIA--PGAQIQHKSIGVSGVALPL 404
Db 344 AFGHGIFECVGAALARLEAQIANGRLRERFPLRMASPDLLRWRFSVLKLEKLPV 401

RESULT 12
CPKX_BACSU STANDARD; PRT; 410 AA.
ID CPKX_BACSU STANDARD; PRT; 410 AA.
AC 008469;
DF 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE CYTOCHROME P450 (EC 1.14.-.-).
GN CYP4 OR CYP107J1.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / BGSJ11;
RX MEDLINE; 97431495.
RA Bialsky B.R., Gustafson M.C.U., Sonenshein A.L., von Wachenfeldt C.;
RT "An lrp-like gene of Bacillus subtilis involved in branched-chain
RT amino acid transport."
RL J. Bacteriol. 179:5448-5457(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE; 97453479.
RA Sorokin A., Bolotin A., Purnelle B., Hilbert H., lauber J.,
RA Duesterhoefte A., Ehrlich S.D.;
RT "Sequence of the Bacillus subtilis genome region in the vicinity of
RT the lev operon reveals two new extracytoplasmic function RNA
RT polymerase sigma factors Sigv and Sigz."
RL Microbiology 143:2939-2943(1997).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Y11043; CAAT1937.1; -
CC DR EMBL; U93876; AAB80898.1; -
CC DR EMBL; Z99117; CAB14615.1; -
CC DR HSSP; 000441; 10XA.
CC DR SUBTILIST; BG11929; CYP4.
CC DR PFAM; PF00067; P450.1.
CC DR PRINTS; PR00359; BP450.
CC DR PROSITE; PS00086; CYTOCHROME_P450.1.
CC Oxidoreductase; Monooxygenase; Heme.
CC KW
```



```

CC or send an email to license@ib-ib.ch).
CC -----
DR EMBL; 277137; CAB00896.1; -.
DR HSSP; P33006; 1CPT.
DR TUBERCULIST; RV1256C; -.
DR PFAM; PF00067; P450; 2.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.
FT BINDING 354 354 HEME (BY SIMILARITY).
SQ SEQUENCE 405 AA; 44580 MW; 72DEAE6CB88FA48 CRC64;

Query Match 14.7%; Score 319.5; DB:1; Length 405;
Best Local Similarity 29.5%; Pred. No. 3.2e-17;
Matches 102; Conservative 51; Mismatches 160; Indels 33; Gaps 9;

QY 62 HWIATRGQILIEAYEDYHFESECPPIREAGE-----AYDFITPSM-DPEQROFRALA 115
   ::::: 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 46 YVLSRHADVMSAANDHQTFFS-AGLVVNGELMIGLHDPRVWMDPRPHTEPKIV 104
   ::::: 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 116 NOVGMPPVYDKLENRIQELACLIESLRPOCCNFTEYAEPPFIRIFMLLAGLPEEDIP 175
   ::::: 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 105 SRGFTPROVEVEPEYPRKFRFVERLEKLRANGSGDITVELFRLPSMVVAHYGVPEEDWT 164
   ::::: 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 176 HUKYLLDQMTNR---DGSMTFA-EAKEALHYLIPIIEORQRKPTDALS-IVANGQVN 229
   ::::: 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 165 QFDGCTQAIIVANAANVADGATGTGALDVAGVSMMAFEGYLIRRKREPDADISHLVAAGVGAD 224
   ::::: 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 230 GRPIYSDAKKMCGLLVGGLDVTYVNFYSFSEMFIAKSPENHROELIERPERIPACEELL 289
   ::::: 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 225 GDVAGTSLIAFTFMVYVGNDYVIGMLGSGMPLHRRPDQRLLDDPEGIPDAVEELL 284
   ::::: 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 290 RRFSLVAD-GRILSYDVEFHGVQLKKGDQILLPOMLSGLDERK-----NACPMHY 338
   ::::: 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 285 RLTSVVOGLIARTTTFEDVYIGDPTTPAGRRVLLLYGSANDERQYGPDAAEILDVTRCPRI 344
   ::::: 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 339 DFRKRVKSHTFEGHSHICLGQHLARRRIITYLAKMFLRIPDFSLA 384
   ::::: 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 345 -----LTFSGAHHCIGAAAAAMOCRAVALTELLARCPDEFVA 381
   ::::: 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111

RESULT 15
CPXL_PSESP STANDARD: PRT: 428 AA.
AC P33006;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CYTOCHROME P450-TERP (EC 1.14.-.-) (CYTOCHROME P450 108).
GN CYP108 OR TERP.
OS Pseudomonas sp.
OC Bacteria; Proteobacteria.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE; 92332528.
RA Peterson J.A., Lu J.-Y., Geisselsoder J., Graham-Lorence S.,
   Carmona C., Withey F., Iorence M.C.;
RT "Cytochrome P-450terp. Isolation and purification of the protein and
   cloning and sequencing of its operon.";
RT J. Biol. Chem. 267:14193-14203(1992).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE; 94166081.
RA Hasemann C.A., Ravichandran K.G., Peterson J.A., Delsenlffer J.;
RT "Crystal structure and refinement of cytochrome P450terp at 2.3-A
   resolution.";
RL J. Mol. Biol. 236:1169-1185(1994).
CC -!- FUNCTION: INVOLVED IN A ALPHA-TERPINEOL OXIDATION SYSTEM.
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its

```

CC used by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M91440; AAA2596.1; -.  
DR PIR: S27653; S27653.  
DR PIR: A42971; A42971.  
DR PDB: 1CPT; 31-JAN-94.  
DR PFAM: PF00067; P450; 1.  
DR PRINTS: PR00359; BP450.  
DR PROSITE: PS00086; CYTOCHROME\_P450; 1.  
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;  
KW 3D-structure  
FT BINDING 377 377 HEME.  
FT TURN 3 4  
FT HELIX 9 16  
FT HELIX 19 22  
FT HELIX 24 37  
FT STRAND 40 43  
FT TURN 46 47  
FT STRAND 51 54  
FT TURN 57 65  
FT HELIX 67 69  
FT TURN 70 71  
FT STRAND 82 91  
FT HELIX 92 94  
FT TURN 94 104  
FT HELIX 102 104  
FT TURN 107 108  
FT HELIX 109 118  
FT TURN 119 120  
FT HELIX 123 145  
FT TURN 146 146  
FT STRAND 150 152  
FT HELIX 153 156  
FT TURN 157 160  
FT HELIX 161 170  
FT TURN 171 171  
FT HELIX 174 185  
FT TURN 186 188  
FT HELIX 208 230  
FT TURN 231 232  
FT HELIX 238 244  
FT STRAND 246 246  
FT STRAND 251 251  
FT HELIX 254 285  
FT HELIX 287 295  
FT HELIX 297 311  
FT STRAND 314 314  
FT STRAND 318 321  
FT STRAND 325 327  
FT STRAND 328 329  
FT TURN 330 332  
FT STRAND 334 335  
FT TURN 337 340  
FT STRAND 342 345  
FT HELIX 346 347  
FT TURN 349 351  
FT TURN 355 356  
FT TURN 360 361  
FT TURN 369 370  
FT TURN 374 375  
FT TURN 378 379  
FT TURN 380 397  
FT HELIX 398 403  
FT STRAND 408 409  
FT STRAND 415 415  
FT STRAND 416 420  
FT STRAND 422 427  
SQ SEQUENCE 428 AA; 47922 MW; 29B772460CC4E93F CRC64;

Query Match

14.68; Score 318; DB 1; Length 428;



Best Local Similarity 26.1%; Pred. NO. 4.4e-17;  
Matches 100; Conservative 61; Mismatches 170; Indels 52; Gaps 9;

[illegible]

Search completed: October 4, 2000, 13:04:52  
Job time: 1683 sec

**THIS PAGE BLANK (USPTO)**